



EJBI 2010

ISSN 1801-5603

European Journal for Biomedical Informatics

Volume 6 (2010), Issue 1

English version

An Official Journal of the European Federation of Medical Informatics

www.ejbi.eu



Content

English version

- en 2 Editorial
Svačina Š., Zvárová J.
- en 3 - 7 The Problems of Astroglial Brain Tumours
Bielníková H., Buzrla P., Dvořáčková J.
- en 8 - 14 Process-based Multi-agent Architecture in Home Care
Bošanský B.
- en 15 - 19 Temporomandibular Joint Prosthesis - 3D-CT Reconstruction before and after Treatment
Feltlová E., Dostálová T., Kašparová M., Daněk J., Hlíňáková P., Hubáček M., Nedoma J.
- en 20 - 26 Methods of the Survival Analysis
Fürstová J.
- en 27 - 32 Electronic Health Record for Temporomandibular Joint Disorders - Support in Therapeutic Process
Hippmann R., Nagy M., Dostálová T., Zvárová J., Seydlová M., Feltlová E.
- en 33 - 37 Comparison of Texture Classifier and Human Observer in Diagnosis of Autoimmune Thyroiditis,
Observer Variability Evaluation
Holinka Š., Smutek D.
- en 38 - 42 On the Relation between Generalized Entropy and the Bayes Decision Error
Horáček M.
- en 43 - 52 Approaches for Constructing Age-Related Reference Intervals and Centile Charts for Fetal Size
Hynek M.
- en 53 - 57 Infrastructure for Data Storage and Computation in Biomedical Research
Kulhánek T.
- en 58 - 65 Language of Czech Medical Reports and Classification Systems in Medicine
Přečková P.
- en 66 - 72 Matching Medical Websites to Medical Guidelines through Clinical Vocabularies in the Context of Website
Quality Assessment
Rak D.
- en 73 - 75 Speeding Up the Algorithm for Finding Optimal Kernel Bandwidth in Spike Train Analysis
Šanda P.
- en 76 - 77 Simple Method of Economical Analysis of Diagnosis Procedure (Used in Screening of Thyroid Gland Diseases
in Pregnant Women)
Telička Z., Jiskra J., Springer D., Límanová Z.
- en 78 - 82 Identification of Units and Other Terms in Czech Medical Records
Zvára K., Kašpar V.
- en 83 - 89 The Database of the Catalogue of Clinical Practice Guidelines Published via Internet in the Czech Language
- The Current State
Zvolský M.

Editorial

It is expected from modern information systems to work effectively with information and to exchange it mutually. If the information is shared using a free text, a prerequisite for semantic interoperability is the access to its meaning. Semantic interoperability and the ability to obtain specific information by usage of technical means are essential conditions for the utilization of other telemedicine technologies and eHealth. The ability of systems to understand exchanged data (semantic interoperability) requires using the same terminology (i.e. classification systems and nomenclatures) and using the same language for communication and its recording (data standards). The utilization of existing norms developed for

our region can ensure the usage of the same terminology. Existing standards (e.g. EN 13606) suppose the use of globally unique and uniquely defined terms that can be without much difficulty transferred to other classifications (e.g. by means of the Unified Medical Language System). Probably the best applicable general classification system for healthcare is SNOMED CT. It has arisen by a combination of American SNOMED (created by the Association of American Pathologists) and British Clinical Terms ("Read Codes"). In connection with this merger the International Health Terminology Standards Development Organisation (IHTSDO), with the residence in Denmark, has been founded.

Within the framework of the specific university research project called *Semantic interoperability in biomedicine and healthcare* the students of the doctoral study at the 1st Faculty of Medicine of Charles University in Prague prepared, as leading authors, the scientific manuscripts on topics of their Ph.D. theses. The papers have been peer-reviewed and now you can find them in this special issue in English and Czech languages.

Prof. MUDr. Štěpán Svačina, DrSc., MBA
Prof. RNDr. Jana Zvárová, DrSc.
editors

The Problems of Astroglial Brain Tumours

H Bielníková¹, P Buzrla¹, J Dvořáčková¹

¹Department of pathology, University hospital Ostrava

Supervisor: Doc. RNDr. Ivan Mazura, CSc.

Summary

Astroglial tumours are a group of tumours derived from the astroglial cells. They belong to a group of most frequently diagnosed tumours of the nervous system. This group of tumours is variable in many aspects. The tumours differ not only morphologically, but also with their growth potential, the extent of invasive growth, their localization within the CNS and the clinical course of the disease due to their progressive features. The variability of these neoplasms is given by the sequence and type of genetic alteration in the course of the neoplastic transformation. According to the currently used classification, it is possible to differentiate between benign I.-II. forms (low grade glioma-I. pilocytic astrocytoma, II. diffuse astrocytoma) and malignant form III.-IV. (high grade glioma-III. anaplastic astrocytoma, IV. glioblastoma) of astroglial tumours. The number of vasoproliferative and necrotic changes, together with the number of cell divisions are the key factors for distinction between benign and malignant tumours from the histological point of view. From the molecular-genetic point of view, it is the types and numbers of cumulating genetic alterations which are of the greatest importance. One type of significant alterations occurs in the p53, PTEN and EGFR genes.

Keywords: astrocytoma, glioblastoma, immunohistochemistry, molecular-biology methods, mutations, tumorsuppressor genes, amplifications, p53 gene, PTEN gene, EGFR amplification

1. Introduction

Astroglial tumours are one of the most frequent primary brain tumours of the intracranial area, responsible for, according to the literature, more than 60 % of primary brain tumours. They are most often classified on the basis of internationally recognized criteria defined by the World Health Organisation (WHO). The tumours may be divided into several

types due to the variability of the histological findings and variable biological behaviour. The astroglial tumours may be localised anywhere within the CNS, they most frequently appear in the brain hemispheres. The growth characteristic of these tumours is mostly infiltrative, and they pose a danger mostly due to their tendency to move towards more malignant forms. Tumours of this group are more often diagnosed in adult patients, some types have been found to have a higher incidence in males [1], [2].

The degree of malignancy and the tumour differentiation may be defined according to a grading system, which is a significant prognostic indicator and is important for the definition of further treatment procedures. The criteria important for the grading determination include proliferation activity of the cells and the degree of cellular differentiation (in the histological picture we follow the maturity of the tumour tissue the degree of differentiation), the presence of cellular and nucleic atypias, angiogenesis and the presence of necrosis [1], [2]. According to their biological features, the astroglial tumours may be differentiated into relatively benign and malignant forms (Tab.1.).

Tab. 1. Grading of astroglial tumours.

Grade	Name of the brain tumour	Tumour characteristics
I	Pilocytic astrocytoma	relatively benign
II	Diffuse (Low-grade) astrocytoma	relatively benign (with a tendency to malignant progression)
III	Anaplastic astrocytoma	malignant
IV	Glioblastoma	highly malignant

To determine the features of the tumour from the histopathology point of view, the

following characteristics are of vital importance: the number of vasoproliferative and necrotic changes, the degree of cellular differentiation, the number of cell divisions and the presence of cellular and nucleic atypias. From the molecular-genetic point of view, it is important to look for the presence of characteristic genetic alterations and their number. Some of the significant changes occur in p53, PTEN, EGFR, Rb and p16 genes.

1.1 The present classification of astroglial tumours

Pilocytic astrocytoma is a Grade I tumour according to the WHO classification. This type of tumour is most frequently observed in children and young adults in various localisations within the CNS, it frequently affects the area of cerebellum and the optic nerve. It is a circumscribed, slowly-growing tumour of rather benign characteristics, with favourable prognosis. The histological picture shows biphasic tumour pattern with areas of densely concentrated GFAP (glial fibrillary acidic protein) positive bipolar astrocytes and eosinophilic Rosenthal threads. Further present are sparsely cellular areas, spumous or even microcystic with GFAP weakly positive protoplasmic astrocytes without fibrils. Regressive changes in the form of hyalinised vessels and calcification are also frequently observed. Nucleic pleomorphias may be also present, however these are less frequent and less prognostically significant. Necrotic changes, angiogenesis or mitosis are not usually observed in pilocytic tumours [1], [2].

The second tumour type, according to the grading, is a diffuse or low-grade astrocytoma. This tumour is usually well-differentiated, slowly growing, however with the features of infiltrative growth and tendencies of malignant progression. It affects especially younger adults. The mitotic activity of the diffuse astrocytoma is low, when compared to higher grades.

Morphologically observed are distinct tumour populations formed by astrocytary cells, which are difficult to differentiate from non-tumour cells. Nevertheless, it is possible to differentiate among three types of low-grade astrocytomas [1], [2].

The most frequent one is fibrillary astrocytoma, where we find nucleic atypias in cells with a low mitotic activity. The typical signs are a formation of microcysts with mucoid contents and the presence of lymphocytary infiltrate in the proximity of vessels. On immunohistochemical demonstration, the cells express GFAP (glial fibrillary acidic protein), S100 and vimentin. The Ki67 proliferation marker shows a low expression.

The second type is gemistocytary astrocytoma. The tumorous cells are usually larger, with plentiful eosinophilic cytoplasm and the presence of perivascular lymphatic infiltrate. The nuclei with small nucleoli are located eccentrically. Immunophenotype and the proliferation activity are the same as with the previous type.

The last type of a low-grade astrocytoma, according to the histological picture, is a less frequently observed protoplasmic astrocytoma. The population of tumour cells is sparse, the cells are small with oval-shaped nuclei and a low content of glial filaments. Formation of microcysts and mucoid degeneration are typical signs of this astrocytoma type, it is possible to focally determine GFAP.

Low-grade astrocytomas are favourable from the prognostic point of view, the average survival period varies between six and eight years. One of the prognostic signs is the Ki67 proliferation activity, when the prognosis may be defined as worse at expression of tumour cells greater than 5 %. The presence of perivascular lymphoid infiltrate and microcysts is a sign of a more favourable prognosis [1], [2].

Anaplastic astrocytoma presents a high proliferation activity and its biological characteristic is malignant. The tumour is diagnosed in adults, more frequently in male population, between the fourth and fifth decade of life. According to the grading it is a III grade tumour. These tumours are

significantly more cellular than the previous types, the cellular nuclei are of various sizes (anisonucleosis), the cell nucleoli are very distinctive. The mitotic activity is elevated, with the presence of typical mitoses. The angiogenesis (formation of new vessels) becomes very apparent, however, in comparison with glioblastomas, the necroses are usually not present. It is possible to determine the expression of S100 protein and vimentin with immunohistochemical determination, the expression of GFAP is variable. The expression of the Ki67 proliferation marker is usually between 5-10 %. The prognosis in this type of tumour is bad, the average survival period varies between two and five years. The tumour may originate from an originally diagnosed diffuse astrocytoma, or may appear *de novo*. The cells lose their differentiation ability and have a tendency towards non-differentiated stages and non-regulated proliferation. The outcome is usually associated with a transfer to the more malignant stage, glioblastoma [1], [2], [3].

Glioblastoma is the most frequently observed type of a malignant brain tumour; it is usually diagnosed in adults between 45 and 75 years of age, with a prevalence in males [2], [4]. We may differentiate between two types of tumours, with regard to the onset mechanism. The primary glioblastoma is more frequent, it appears *de novo*, without any prior tumorous lesions. The onset and clinical course of the disease is usually very rapid. The second type, a secondary glioblastoma, develops gradually, through a malignant progression from anaplastic astrocytoma of another type of a glial tumour. The glioblastomas vary greatly histologically and topographically. Glioblastoma is a Grade IV astrocytoma, the cells show a low degree of maturity with manifestations of anaplasia, the tumour is highly cellular, formed by non-differentiated pleomorphic elements. Extra large cells with large nuclei and typical mitoses are present. We may typically observe necrotic foci (local and massive), especially in the centre of the tumour, where the tumorous cells form typical palisade structures at the periphery of the necrosis. Formation of new blood vessels is vast and plentiful. The immunohistochemical markers are vimentin and CD99, the expression of

GFAP is variable. The proliferation marker Ki67 manifests a positive expression in approximately 15 to 20 % [1], [2], [3], [4].

The average period of survival in patients with a diagnosed glioblastoma is one year from the diagnosis. From the prognostic point of view, the age of the patients, the proliferation activity of the tumour and the level of EGFR expression are significant. The higher the age, the proliferation activity and the expression of EGFR, the worse the prognosis [1], [2], [3], [4].

1.2 Genetic alteration of the astroglial tumours

Thanks to molecular-biological methods, a number of genetic alterations have been defined in the astroglial tumours. These are related to the onset and development of these tumours. The onset of the tumour is usually preceded by an initiation damage of a gene, together with further accumulating defects which lead to the progression of the tumour.

The most frequent initiation point in the astrocytary tumours is a mutation of the p53 gene. This gene is localised on a short arm of the chromosome 17 (17p13.). The defects of this gene are the most frequently observed in human tumours. In most cases mutations of both alleles of the gene are present, however cases of a hereditary mutations of one allele are also described in the literature. The p53 gene produces a transcription factor which, under normal circumstances, stimulates expression of other genes. It is an important tumour suppressor gene participating upon the regulation of the cellular cycle and playing an important role in cellular stress. It blocks the cellular cycle and gives the cell the time to assess the degree of damage and possible subsequent DNA repair. In case the defects cannot be repaired, the p53 gene initiates the mechanisms of apoptosis. The defects of this gene, together with an abundant expression of PDGF (Platelet Derived Growth Factor) are considered to have a connection with an onset of diffuse astrocytomas, which may further progress into anaplastic astrocytoma, or even secondary glioblastoma [2], [3], [6].

The progression of astrocytoma towards a higher degree of malignity is further caused by a loss of the Rb gene. The retinoblastoma gene (Rb) is localised on the long arm of the chromosome 13 (13q14). The product of the gene is a nucleic phosphoprotein, which, under normal circumstances, controls the transcription of other genes participating upon the regulation of cellular division. As in the case of p53, it possesses a tumour-suppressor function [2], [3].

The pilocytic astrocytoma is a Grade I tumour with a relatively benign features which develops through an independent genetic pathway after the loss of the NF1 gene. The NF1 gene is localised on the long arm of chromosome 17 (17q11). The product of the gene is protein-neurofibromin which is a negative regulator of the p21^{ras} protein (it plays a significant role in the process of cellular communication and signalling) [2], [3], [6].

In cases of primary and secondary glioblastomas, the defects of genes are different. A primary glioblastoma originating *de novo* has a separate evolution path, with a direct transformation into highly malignant phenotype without interstages. The specific defects of the primary glioblastoma are deletion of the p16 gene, mutation of the PTEN gene, loss of heterozygosity on the chromosome 10, mutation and amplification of the gene for the receptor of epidermal growth factor (EGFR) and MDM2 amplification [1], [2], [3]. The p16 gene codes the inhibitor of cyclin-dependant kinases 4 and 6 and is located on the short arm of the chromosome 9 (9p21). The suppressor gene PTEN is localised on the long arm of the chromosome 10 (10q23.3). Its functions are the blockade of the cellular cycle in the G1 stage. It is especially disorders of this gene which are linked to the progression of lower-stage astrocytomas into more malignant forms. The presence of such mutation is connected with a worse prognosis of the tumour [2], [3], [6], [7], [8], [9], [10]. The gene for EGFR receptor lies on the short arm of the chromosome 7 (7p12-14) and is the most frequently amplified oncogene in primary glioblastomas. Its amplification most probably plays a role in diffusion

infiltration of glioblastomas into the surrounding tissues. The high numbers of amplifications are linked with a worse tumour prognosis [2], [3], [4], [5], [6], [11].

Another amplified gene MDM2 is localised on the long arm of the chromosome 12 (12q13-14). This gene is responsible for coding of protein, the function of which is the negative regulation of apoptosis through p53 and Rb proteins [6].

In cases of the secondary glioblastoma which develops through a diffuse astrocytoma and anaplastic astrocytoma, the detected mutations are identical with the mutations in lower stages of these tumours. Less frequently, we may also observe amplification of the gene for EGFR receptor, deletion of the p16 gene or mutations in the PTEN gene. Identically with the primary glioblastoma, the loss of heterozygosity on the chromosome 10 is present to a large extent [2], [3], [6].

2. Current diagnostic approaches

The brain tumorous tissue is usually harvested for bioptic examination by the clinician during the harvesting procedure, with the use of a cannula. The acquired material may be submitted for histology examination in a negative form (being sent to the pathologist immediately after harvesting, placed in a moist chamber closed box with a piece of gauze sodden in the physiological solution), or in a fixation solution. A correct harvesting procedure is an important factor for the diagnostics, in cases of an incorrect localisation, a non-tumorous tissue may be harvested. The consequences of such a mistake include expensive, not necessary and time consuming examinations being performed. Another critical step is the processing of the material, with the maximum effort to prevent depreciation of the harvested tissue and prevention of artefacts.

The molecular-biological methods enable us to examine a "fresh" tissue harvested during a biological harvesting procedure, as well as tissue already processed with the method of paraffin blocks.

The main diagnostic methods for the determination of astroglial tumours are histological and immunohistochemical examinations. The histological exami-

nation gives us a morphological picture of the tumour (Fig. 1.),

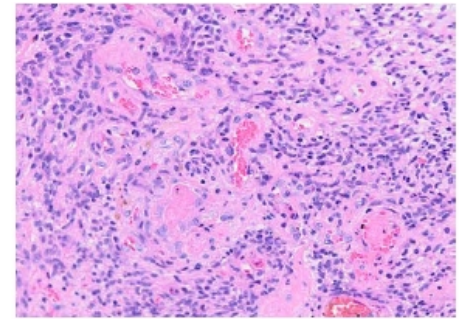


Fig. 1. Histology of glioblastoma (HE, 400x, Department of pathology, University hospital Ostrava).

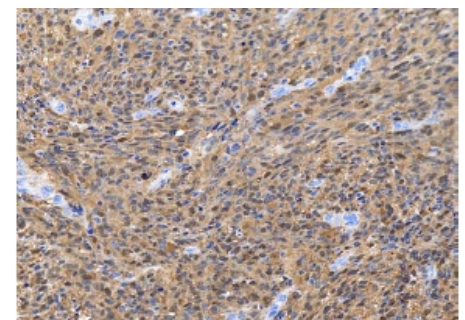


Fig. 2. Expression of S100 protein in glioblastoma tissue (S100, 200x, Department of pathology, University hospital Ostrava).

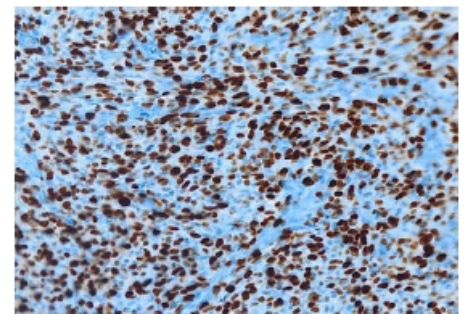


Fig. 3. Expression of Ki67 marker in glioblastoma tissue (Ki67, 200x, Department of pathology, University hospital Ostrava).

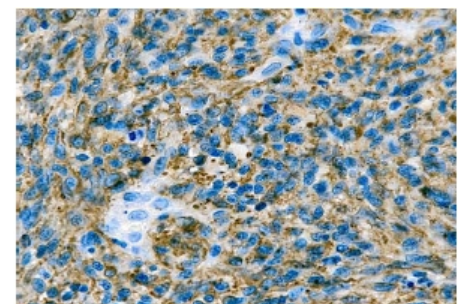


Fig. 4. Expression of GFAP protein in glioblastoma tissue (GFAP, 400x, Department of pathology, University hospital Ostrava).

immunohistochemical examination is a complementary examination, however presents an inevitable part of the bioptic diagnostics.

The most commonly used markers in the problems of astroglial tumours are vimentin, S100 protein (Fig. 2), glial fibrillary acidic protein-GFAP (Fig. 3) and proliferation marker Ki67 (Fig. 4). The markers of connective tissue, vimentin and S100 protein are the proteins of intermediary filaments. Vimentin is expressed in embryonic cells and tissues during the evolution, and in adult mesenchymal cells. It is a predecessor of glial fibrillary acidic protein (GFAP) in immature glial elements and, together with GFAP, is co-expressed in mature astrocytes and their tumorous derivatives. S100 protein is an acidic protein of the central and peripheral nervous system, however it cannot be considered as a specific protein of the nervous tissue as its presence has been shown also in other tissues [12].

Ki67 is a marker of the speed of cellular proliferation. It reacts with the human nucleic antigene expressed in proliferating cells of all stages of the cellular cycle, except the G_0 stage [12]. Immunohistochemical methods enable to carry out the basic typization of tissue and its derivatives.

Together with the discovery and evolution of new methods, namely molecular-biological ones, the possibilities of diagnostics have been moved even further. These new methods present another supplement in the diagnostic process and enable us to obtain further information important for a more specific and precise definition of the disease. The methods can be used for the analysis of known mutations and alterations in the genes the damage of which is known to be related with an onset and development of the tumour. Assessment of these mutations becomes an important prognostic factor which is very important for determination of the further course of the disease and for possible initiation of appropriate and adequate treatment. At the same time, the methods make possible the discovery of new, presently unknown mutations, alterations and other factors,

which may participate upon the onset of the disease. The most frequently used molecular-biological methods include hybridization, electrophoresis, amplification methods (PCR- polymerase chain reaction), sequencing, microarray analyses, etcetera.

3. Possibilities of evidence and processing of the acquired data

Realization of every research project or a clinical trial is connected with a production of a vast amount of information; this information needs to be processed in some way. In case of astroglial tumours we may obtain information related to the course of the disease, genetic and histological data, etc. However, as this disease is always closely linked to a particular person, we also receive a large amount of personal data regarding this patient. We may use database systems to process this type of data, as these enable us to generate various outcomes.

By the term “database” we mean an exactly defined data file used to describe a real world. In the case of astroglial tumours such a file may contain a record of patients with astroglial tumours, where the patient presents the element from the real world, so called entity, and this element is further described by its specific features - *attributes* [13], [14]. The usually used attributes are name and surname, residence address, birth number, etc. In relation to the disease, such attributes may be the type of the tumour, the degree of grading, detected mutations, and so on.

One of the means of describing a database is a database model. The most frequently used models are relational databases, which may be imagined as a system of tables consisting of lines and columns, where the columns correspond with individual attributes (e.g. name, surname, residence address, etc.) and the lines describe the current state of the patient. In that respect, a very important part of the database model is formed by a so called primary key, the feature of which is having a unique value. The unique value of the primary key prevents a situation when for example two lines in the table could have an identical value of the primary key, which

is an unwanted situation. Such a key in case of patients may be presented by the patient's birth number. The problem is, however, that, in relation to the Law concerning the protection of personal data, the birth number has been defined as a classified information and as such it cannot be used. In case we are not able to use a certain column of the table with data as the primary key, or possibly a combination of columns, it is usually possible to add another, so called “id (identification) column”, which is a column with unique values of gradually rising numbers. It is also possible to create a primary key by compounding the values of all columns, however such a solution is not very suitable, due to the fact that it increases the computing requirements of the database system. That is why so called “normal forms” have been defined for the database systems, which is a list of rules according to which the database should be defined for an optimal use of the database system. [13]. The existing relations among the tables with certain columns create certain connections which may be found also in the real world. For example a patient (as a table with the columns titled name, surname, birth number, name of the tumour) has a large tumour diagnosed. This relation is defined in the *tumour* column from the table *patient* with a relation to the primary key of the column *name of the tumour* from the table types of tumours with concrete values *large tumour*, *small tumour*, *little tumours*.

The most frequently used language for a formation of databases in the world of information technologies is the SQL (Structured Query Language) and the database systems based upon this language. The language presents a tool for formation of databases (tables) and also enables the users a further manipulation with data, like data storage, update, erasing, searching for specific information, etc. [14].

When defining a database, it is necessary to evaluate every step because every mistake or a wrong step may be critical and these mistakes may cause a malfunction of the database in the future, during transferring and editing the stored data, etc.

Another possibility of editing the acquired data is the use of statistical methods. There exists a wide range of these methods, the choice of the appropriate method needs to be based upon the type of data which we need to process.

The most frequently observed topic in relation to the astroglial tumours is the effectiveness of prognostic factors on the survival of patients with glioblastomas. To assess the mean survival period in patients we use the Kaplan-Meier method and the effect of the variable on the survival is evaluated with the Log-Rank test. The variable is often the age of the patient, sex, Karnofsky score, localization of the tumour in brain, postoperative radiotherapy, chemotherapy and others [15]. Other statistical methods frequently used in the area of astroglial tumours are Chi-Quadrant-Test, Student's t-test, Mann-Whitney U test or Spearman's correlation coefficient [16].

4. Conclusion

The problems of astroglial tumours are very vast and complex, even though a large number of clinical trials have been conducted in that respect, many questions remain unanswered. The most important point is looking for new prognostic and predictive factors which would have a significance for choosing suitable therapeutic procedures. This problem is mostly related to the glioblastomas, where the period of survival from the time of diagnosis is very short. From the overall number diagnosed glioblastomas, around 90% are primary ones. This type of tumour develops very quickly, without any previous clinical or histological manifestation. The primary and secondary glioblastomas vary, apart from different mutations, also in the profiles of expression on the RNA and protein levels in the methylations of promoter areas.

These differences are responsible for a different susceptibility to radio- and chemotherapy [17], [18]. A large percentage of the present clinical trials are aimed at the use of EGFR receptor as the molecular target in cancer therapy. It has also been proved that methylation of O⁶-Methylguanine DNA methyltransferase (MGMT), a key DNA-reparatory enzyme, increase the sensitivity of secondary glioblastomas to alkylating agents used in chemotherapeutic treatment [18], [19].

Obtaining new and more specific information regarding the pathogenesis of astroglial tumours which would result in an improved and more quality diagnostics and treatment of these tumours is a long-term task which will require a willing cooperation among clinicians, pathologists and molecular biologists.

Acknowledgment

The paper was supported by the project SVV-2010-265513.

References

- [1] Ehrmann J., Koďoušek R., Kolář Z.: *Stručný přehled patologie nádorů nervového systému*, LFUP Olomouc; 2007.
- [2] Louis D.N., Ohgaki H., Wiestler O.D., Cavenee W. K.: *WHO classification of tumours of the central nervous system*. Inter. Agen. for Res. on Can.; 2007.
- [3] Coleman W.B., Tsongalis G.J.: *Molecular pathology; The molecular basis of human disease*. Elsevier; 2009.
- [4] Ohgaki H. et al.: Genetic pathways to primary and secondary glioblastoma. *Amer. Journ. of Pathol.* 2007; 170: 1445-1453.
- [5] Puget S. et al.: Malignant brain tumors: Two steps forward. *Clin. Neurosur.* 2007; 54: 4-9.
- [6] Kolář Z. a kolektiv: *Molekulární patologie nádorů*. EPAVA Olomouc; 2003.
- [7] Duerr M.E. et al.: PTEN mutations in gliomas and glioneuronal tumors. *Oncog.* 1998; 16: 2259-2264.
- [8] Aghi K.M. et al.: Decreased rate of infection in glioblastoma patients with allelic loss of chromosome 10q. *J. Neurooncol.* 2009; 93:

115-120.

- [9] Rasheed Ahmed B.K., Stenzel T.T., et al.: PTEN gene mutations are seen in high-grade but not in low-grade gliomas. *Canc. Res.* 1997; 57: 4187-4190.
- [10] Wang I. et al.: Somatic mutations of PTEN in glioblastoma multiforme. *Canc. Res.* 1997; 57: 4183-4186.
- [11] Frederick L. et al.: Diversity and frequency of epidermal growth factor receptor mutations in human glioblastomas. *Cancer Research* 2000;60: 1383-1387.
- [12] Lukáš Z., Dráberová E. a kol.: *Imunohistochemické metody v biologii a v biotické diagnostice*. MU Brno; 1997.
- [13] <http://interval.cz/clanky/database-a-jazyk-sql/>
- [14] Oppel A.: *SQL bez předchozích znalostí*, Comp. Press. 2008; 240.
- [15] Tugcu B., Postalci L. S., Gunaldi O. et al.: Efficacy of clinical prognostic factors on survival in patients with glioblastoma. *Turk. Neurosur.* 2010; 2: 117-125.
- [16] Martinez R., Rohde V., Schackert G.: Different molecular patterns in glioblastoma multiforme subtypes upon recurrence. *J. Neurooncol.* 2010; 96: 321-329.
- [17] Voldborg R.B. et al.: Epidermal growth factor receptor (EGFR) and EGFR mutations, function and possible role in clinical trials. *A. of Onc.* 1997; 8: 1197-1206.
- [18] Kuan C-T, Wikstrand C.J., Bigner D.D.: EGF mutant receptor vIII as a molecular target in cancer therapy. *Endocr.-Relat. Can.* 2001; 8: 83-96.
- [19] Eoli M., Menghi F. et al.: Methylation of O⁶- methylguanine DNA methyltransferase and loss of heterozygosity on 19q and/or 17p are overlapping features of secondary glioblastomas with prolonged survival. *Clin. Can. Res.* 2007; 13: 2606-2613.

Contact

Mgr. Hana Bielníková

Department of Pathology
University Hospital Ostrava
17. listopadu 1790
708 52 Ostrava-Poruba
Czech Republic
e-mail: hana.bielnikova@email.cz

Process-based Multi-agent Architecture in Home Care

B. Božanský¹

¹Dept. Of Cybernetics, Faculty of Electrical Engineering, Czech Technical University in Prague

Supervisor: Doc. Ing. Lenka Lhotská, Csc.

Summary

Utilization of procedural knowledge in the form of organizational processes and formalized medical guidelines can be useful in decision support systems (DSSs) in health care domain. The problem of using this form of knowledge arises when a multi-agent paradigm is to be applied in a DSS due to differences in specification of behavioural models of agents and process formalisms. In this work we continue in enhancing a novel process-based multi-agent architecture and demonstrate its integration into an existing DSS (K4care) focused on home care.

We analysed available documentation of the complex system K4Care and identified possible mutual common functionalities of implemented multi-agent system with the new architecture. These were the entry points, using which we further enhanced the K4Care platform with respect to the process-based multi-agent architecture.

The analysis proved that the integration is not only possible, but thanks to the general design of the process-based multi-agent architecture can be done with only small changes in the existing K4Care model. Immediate improvements in supporting human experts were identified and possible further improvements of the system were discussed.

Adopting the process-based multi-agent architecture can be beneficial even for existing DSSs and can open new possible features emerging from the multi-agent paradigm.

Keywords: multi-agent systems, multi-agent architecture, organizational processes, formalized medical guidelines, health care, home care, K4Care

1. Introduction

Decision support systems (DSSs) have established their place in the field of health care and assist physicians with their decisions. Several studies proved that

decision support systems (DSSs) can be useful for preventing errors, or improving the quality of provided care, they can reduce costs and save time [1-5]. The decision support systems can be helpful in practice only when they offer unobtrusive interface to users, such as physicians and/or other health care specialists. This observation is proved in a systematic review of existing DSSs [2], where systems indicated as most helpful were working transparently in the background and automatically prompted the user in case of a reminder or an alert (e.g. systems [6], [7]). The knowledge used by DSSs is usually based on medical guidelines (e.g. in systems [8], [9]) that describe a recommended procedures for diagnosing specific diseases and/or their treatment. However, based on the analysis of existing DSSs [1-9] we conclude that a successful DSS is fully integrated into a medical facility where it is used in order to fit current circumstances. Therefore organizational processes must be considered when creating the procedural knowledge for a DSS. Such a procedural knowledge can then be used by a DSS for process validation, simulation or verification of their execution in practice.

One of the fields of health care where applying DSSs can be highly beneficial, is the domain of Home Care (HC) [10], [11]. Growing number of chronic elderly patients result in growing need for a lifelong treatment under continuous health care specialist supervision. Moreover placing the patients at their home environment can have better results compared to improper effects of institutionalization [11]. As the home care domain is distributed in nature, there is a need for appropriate DSS. Health care specialists visit patients at home in order to provide appropriate assistance, however, they cannot easily consult specific questions related to unfamiliar situations or complications. Therefore,

a DSS supporting and monitoring their actions can improve provided care.

The distributed nature of the home care have to be reflected by a DSS as well, therefore a multi-agent paradigm is often used [11-15]. As a multi-agent system we term a system where a group of autonomous intelligent agents jointly interact in order to achieve predefined goals. The behaviour of an agent can be specified using various behavioural models, such as (1) purely reactive models where the agent only reacts to the changes present in the environment, (2) deliberative models, where the agent create its own model of the environment and reason about its future actions, and (3) hybrid models, combining the two previous models. These behavioural models can be formalised using general programming languages, such as Java, or they can be formalised using some agent programming languages, such as Jazzyk [16], JASON [17], or versions of the APL language [18]. The problem of existing agent programming languages is that they offer only limited capabilities of effective modelling of implicit cooperation and coordination between agents in order to make them concurrently follow both private and common goals that require cooperation. On the other hand, organizational processes and medical guidelines are highly based on cooperation of involved actors and provide the information about the order and conditions for execution of specific activities (e.g. perform an examination prior to a surgery, which also need specific equipment, etc.). Therefore in the previous work we defined a new way of creating the behavioural models for intelligent agents based on a novel multi-agent architecture that would directly support such cooperation [19], [20].

The approach is based on enhanced processes, in which a procedural knowledge can be formalised. Based on this information we create a set of agents in a specific architecture that are able to work with these processes in terms of their validation, simulation, or verification of their execution and critiquing in practice. For explanatory reasons we refer to this architecture as ProMA in this work.

The developed architecture (ProMA) is designed to work with general process-like formalization. Hence the goal of this paper is to demonstrate a possible application of ProMA for extending an existing architecture for a home care system K4Care [11], where the formalism SDA* [21] is used for storing the procedural knowledge. We analyze the impact of an integration of a new multi-agent architecture into the K4Care system in terms of necessary model changes and new features. The paper is organized as follows. In the next section we introduce the concept of processes in health care, following by the description of the ProMA architecture that is able to work with them. Next section is devoted to the brief description of the K4Care system, used architecture and functioning of the system. Following sections describe the integration of the ProMA architecture into the K4Care system. This integration acts as a proof of concept of applicability of ProMA. Finally, we propose possible new features for the K4Care system based on the integration, and we conclude our approach in the final section of the paper.

2. Process-based Multi-agent Architecture (ProMA)

In this section we define processes in health care, identify two basic types of the processes, and describe the ProMA architecture that is able to work with them. We understand the term process to be a sequence of activities, states, decision points, steps splitting the sequence, and synchronization points. Processes in health care can be divided into two main categories: organizational processes and medical treatment processes (medical guidelines).

The organizational processes in the health care domain are closely related to other

business areas. There are several studies (e.g. in [22], [23]) that practically examine problems of applying process modelling or workflow management systems within the health care domain. There is a mutual agreement that a successful implementation of this approach can improve the quality of health care, reduce the time of necessary hospitalization and lead to the reduction of costs.

Medical guidelines are for some time now a part of the process of standardization of medical treatment and act as pillars of evidence-based medicine. Guidelines contain recommended actions, directions, and principles for diagnosis or treatment specific diseases approved by appropriate expert committees. However, the work-methods regarding the guidelines (such as consulting or using in practice) are still usually based on a textual form, which is natural for health care specialists. The textual form is, however, not sufficient for a quick consultation during an examination of a patient, or an identification relevant changes in the new version of the document. Moreover, guidelines in the textual form cannot be directly used for computer interpretation, or as a knowledge base for decision support systems. Therefore a wide part of research in biomedical informatics is related to the formalization of medical guidelines into an electronic form and various process

languages are used in order to capture the knowledge of a textual medical guideline into an electronic and structured form (e.g. PROforma [24], Asbru [25], GLIF [26], etc.).

Both of these types can be jointly seen as processes in health care and have to be reflected in the procedural knowledge base of a DSS. In following we describe the architecture of a multi-agent system that is able to work with procedural knowledge captured using a generic process-like formalism. The scheme of the architecture is depicted in Figure 1.

2.1 Environment Agent

An agent-based simulation is situated in some environment that is represented by the Environment Agent in the ProMA architecture. With respect to the level of detail that we want to model using the system, the environment could represent a virtual world (e.g. database of patients' health records, a whole department of a hospital, etc.) with existing objects (e.g. laboratory results, RTG or EEG machines, beds, etc.). There is one difference between handling data from patient's electronic health record (EHR) and other objects in the environment. As it is shown in the figure, the agent managing the patients' EHR is requested by the Environment Agent as patients' data certainly are a part of environment.

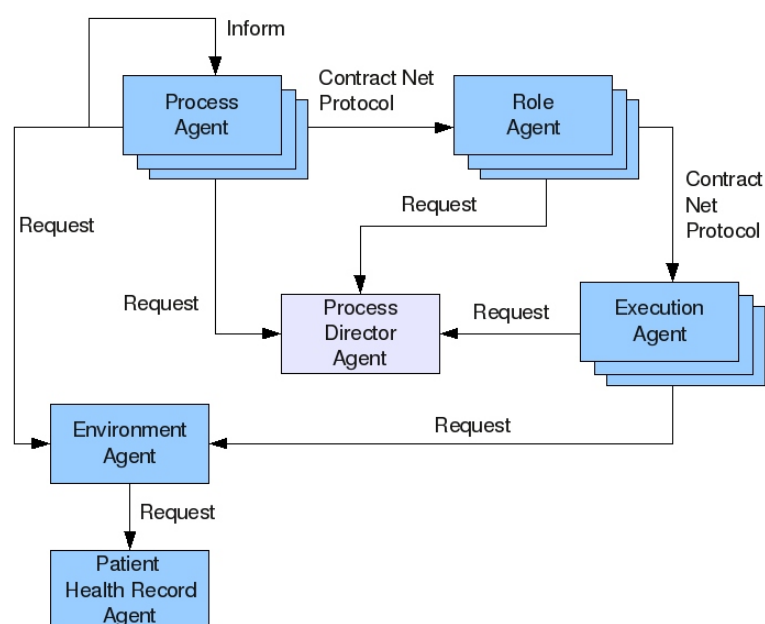


Fig. 1. Scheme of the process-based multi-agent architecture.

However, as it would be described in the next sections, there could be a need for a distinction between real data stored in the record (Patient Health Record Agent) and data used for a simulation (EnvironmentAgent).

2.2 Execution Agent

Execution Agents (EA) represent concrete physicians, nurses, patients, and/or possibly other employees of a facility involved in the processes. These agents are based on a reactive architecture in the form of hierarchical rules, which can be automatically generated based on possible activities that the specific EA can participate in (see [20] for details on the transformation algorithm). Each EA has several predefined rules that cover basic behaviour in the environment (i.e. responding to messages sent by other agents, sending appropriate messages to the Environment Agent during the execution of the activity, etc.). Then, for each activity that the agent (hence the represented person) can participate in, one additional rule is generated. These rules can be activated (when the conditions for the process execution are met, and the EA can perform this action) or deactivated (execution of this process is no longer possible) by a message sent by appropriate a Role Agent (see below). Finally, the Execution Agent autonomously chooses which of the activated processes it executes based on its internal priority.

2.3 Role Agent

Role Agents (RAs) represent the roles in the environment (i.e. general roles for patient, nurse, physician, etc.). RA receives a proposal from a Process Agent (see below) and finds appropriate Execution Agent(s) (EA). Special agents for roles reflect the typical hierarchical structure of roles at a workplace (e.g. a secretary, a nurse, or a doctor are all also employees, etc.). Therefore, when a RA receives a proposal from a Process Agent, it starts to find the appropriate EA between agents that possess this role (using a contract-net protocol (CNP)), but also roles, that are more general in the hierarchy.

If multiple EAs are able to execute given activity and only one is needed, the RA will select the best of them according to its

internal rules and send the message activating appropriate rule to the selected EA. The method of selecting an EA is always domain or role dependent (e.g. the EA that is idle for the longest time can be chosen, etc.). Moreover, RAs are responsible for finding another EA in case the current one needs to suspend its participation on currently executed activity (e.g. one physician is needed in a case with a higher importance and he/she can be replaced in the current one, etc.).

2.4 Process Agent

For every step in the process notation (i.e. for each activity, event, decision point, etc.) there is one Process Agent (PA) created in the system. PA(s) are responsible for a proper execution of the associated step. At first, a Process Agent checks whether the initial conditions for the process are met:

- whether the previous PA(s) has successfully finished its execution (i.e. whether the PA has received the inform message from its predecessor(s)),
- whether all input objects have correct values (the PA uses a simple request protocol to retrieve these data from the EnvironmentAgent),
- whether there exist appropriate agents that will execute this action (the PA uses a CNP for those RAs that are connected to this activity).

If all mandatory conditions hold, the PA starts the execution of the step (e.g. simulation of an activity, a calculation, or a decision process, etc.) and after a successful finish, the PA is responsible for notifying the Environment Agent about results of the activity (using simple request protocol) and the next succeeding Process Agent about a successful finish (using simple inform protocol).

Our approach takes into account a possibility of a temporal suspension of activities and reflecting partial results in the environment, replacing executing EA with another, coordination of several EAs participating on a single activity, or optional input conditions.

2.5 Process Director Agent

The last agent is an auxiliary agent that reads formalized processes, creates other

agents, and answers them to their requests regarding information about the processes (e.g. input data, predecessors, etc.). Note that this is only an implementation issue and all agents reason in terms of processes.

3. K4Care

The aim of the K4Care project [11] was to create, implement, and validate a knowledge-based health care model for the professional assistance to senior patients at home. The procedural knowledge is represented by clinical guidelines, termed within this project as Formal Intervention Plans (FIPs). However, in order to capture specific conditions, in which each patient is present, the K4Care project introduces Individual Intervention Plans (IIPs) that merge both medical and social characteristics of the patient with the available medical knowledge that is represented by the FIPs.

All procedural knowledge related to both types of Intervention Plans (Formal and Individual) is represented in K4Care using the SDA* language [21]. The language has a process-like structure, in which there are three main components (called terms): (1) state terms, (2) decision terms, and (3) action terms. State terms are used to describe feasible patient conditions or situations in the area of interest (e.g., terms as high-blood-pressure are used to emphasize the state). Decision terms represent the terminology that health care professionals use to differentiate possible alternatives (e.g. term antecedents-of-heart-problems would derive the course another direction if it is true). Finally, the action terms identify the medical, clinical or management activities performed by actors (e.g., actions like take-beta-blocker, avoid-salt-in-meals, make-blood-analysis, etc.). Compared to typical process-like formalisms, the actors, as a specific component, are missing in the SDA* formalism. There is, however, a fixed set of actors assigned to each action. More specifically there are two sets of actors - the set of *petitioners* and the set of *performers*. Actors in the first set are allowed to request the related action to be executed. Actors in the set of performers are those persons that are allowed to execute the action.

Formalized procedural knowledge is used by human experts (physicians, or medical centre managers) at first. In order to provide a new service to a request made by a patient registered in the system, a team of specialists identify possible diseases based on the signs and symptoms of the patient. The platform automatically retrieves a set of Formal Intervention Plans (FIPs) that considers all of identified factors. All of these FIPs are merged together and modified by a human expert in order to cover specific circumstances creating this way an Individual Intervention Plan (IIP). In an IIP some other actors can be involved, hence their approval is necessary. After the approvals of the procedure, the responsibilities resulting from the IIP are generated by the SDA* execution engine and specific tasks are generated for involved actors. After a health care specialist executes specified task and he/she fills a necessary data into a document related to the action, and another task is generated according to the IIP.

4. Integration of ProMA into the K4Care System

In this section we provide a description of an integration of the existing K4Care model and the ProMA architecture.

We describe necessary conceptual and technical changes at first, following by a proposition of further capabilities of the modified K4Care system. In this section we follow the typical use-case of the K4Care system described at the end of the previous section and illustrate the integration within this context. An immediate advantage of implementing the ProMA architecture into the K4Care system is in improved decision support for a human expert creating a new IIP for a patient. However, further possible improvements are available as we discuss in the next section.

4.1 Environment

We are focused on the multi-agent system part in the K4Care architecture (see Figure 2). First of all, we unify existing functionality of the K4Care system with ProMA.

The connection to the health record data and procedural knowledge of the multi-agent system represented in ProMA as the Patient Health Record Agent and the Process Director Agent is in K4Care realized by the Data Abstraction Layer (DAL). This layer provides Application Program Interface (API) for requests about patient's data from the health record or guidelines formalized in SDA*.

However, the Environment Agent must still be present in order to temporarily store simulation-based data (that are not persistent). The DAL in the current version is not able to satisfy partial request on the structure of the formalized processes as initialized by Process or Role Agents in ProMA (DAL was designed only for retrieving complete documents). Therefore, both, the Process Director Agent and Patient Health Record Agent, should be created in the multi-agent system as well in order to respond to these requests.

4.2 Execution Agents

The Execution Agents (EAs) in ProMA conceptually correspond to Actor Agents in the K4Care system - they represent the actual users of the system and executors of the actions. In order to fit the ProMA model completely, Actor Agents must be able to communicate with Role Agents (RAs), and autonomously react on their requests for participation in an activity. Note, that in the current version of the K4Care system providing this response is solely based on the activity of human users.

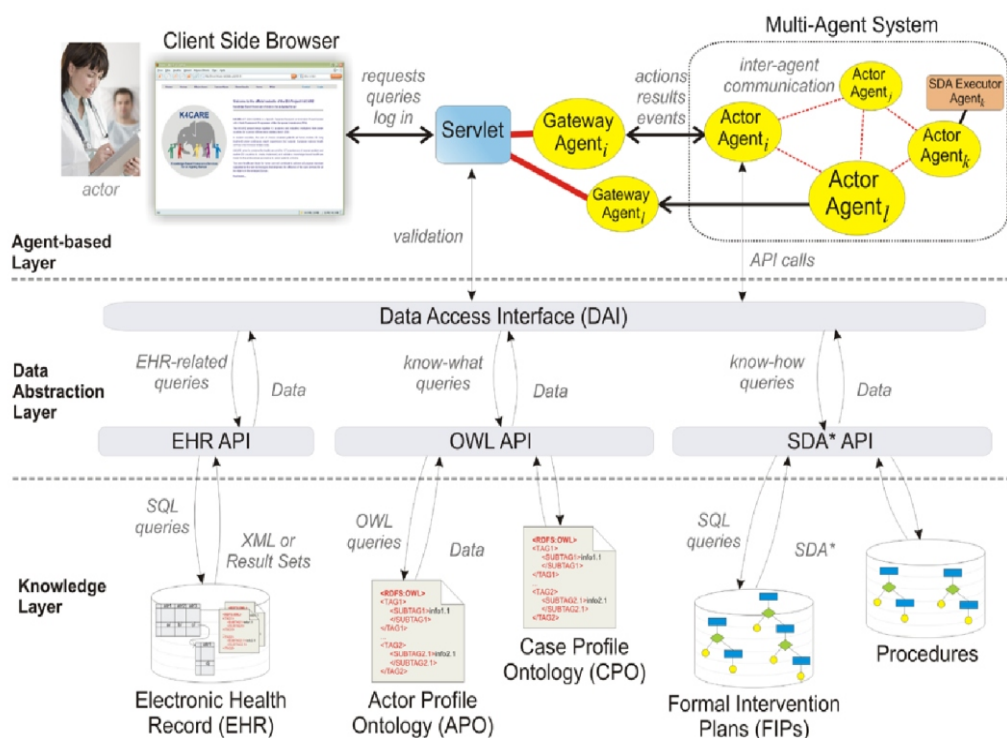


Fig. 2. Scheme of the K4Care system as presented in [13].

However, due to the persistent presence of each Actor Agent in the system, their functionality can easily be enhanced with the ability to respond to the contract-net protocol (CNP) initiated by Role Agents (RAs). The respond can be based on currently associated tasks and/or availability of the associated user (e.g. if a medical specialist has left the facility for a longer time, the Actor Agent could possess this information and automatically deny his/her participation on proposed activities). Moreover, the SDA* concepts typically have the notion of time-related information stored (i.e. time interval and step of the occurrence for a state or an activity), hence Actor Agents can provide much more sophisticated answers based on currently planned actions and the associated IIPs, and act as a basic time-scheduling algorithm easing a human expert when creating a new IIP for a patient (see the next section for new possible features).

4.3 Role Agents

The rest of the types of agents presented in the ProMA architecture do not have their alternative in the K4Care model, hence they have to be created. In the following we describe their full integration into the K4Care system. Note that for Role Agents (RAs), neither Process Agents (PAs) is necessary to be persistent in the system within proposed use-case of creating an IIP (only relevant roles and processes should be created based on requested FIPs).

Role Agents (RAs) conceptually correspond to types of actors in the K4Care system. For each actor type we create one RA as follows: Patient, Family Doctor, Head Nurse, Physician in Charge, Social Worker, Nurse, Specialist Physician, Social Operator, Continuous Care Provider, and Informal Care Giver. Moreover, some of these types belong in the K4Care system into groups Evaluation Unit and Additional Care Givers. Therefore for each of these groups one RA is created as well and a hierarchy between the RAs is established. Moreover, further hierarchy can be seen in various activities (most of the activities of a nurse can be executed by a head nurse as well, etc.).

4.4 Process Agents

Finally, for each state, decision or an activity in requested FIP SDA* models one Process Agent (PA) is created in the multi-agent system. When the IIP is being created by the user, for each step one PA is created as well. The collective behaviour of the PAs during the execution of the IIP by health care specialists correspond to the behaviour implemented the SDA* Execution Engine. However, the advantages of using PAs for each step in the SDA* procedure is during the process of creation of a new IIP by merging the requested FIPs. This process can be partially automated, hence offer some preliminary results that a human expert can use and further improve according to his/her knowledge.

All of the characteristics of the SDA* steps in the K4Care system are formalized using an exhaustive ontology structure, hence we can utilize these information and try to automatically merge some of the steps from requested FIPs in order to simplify the final model. The equivalence of several steps can be identified by appropriate Process Agents (PAs) comparing associated terms and context. If they match, a single PA can be created in the IIP with the union of needed documents and output data (e.g. a blood-analysis can be performed only once with a union of desired information). Although the idea of semi-automated merging whole SDA* procedures was briefly introduced in [13], we propose merging of separate steps of the procedures using autonomous agents. This can be further improved by using a basic simulation (will be described below) of all of the processes concurrently (which ProMA offers). The PAs can identify the order in which they can be executed based on fulfilment of their prerequisites. By the term "basic simulation" we mean a simulation, where the PAs try to simulate associated steps by temporarily filling necessary data fields of the patients' data profile with varying arbitrary values and analyse which of following PAs have fulfilled their prerequisites. We further improve this insight in the next section by more advanced simulation.

Finally, using the CNP Protocol, Process Agents (PAs) can delegate a problem of selecting appropriate health care specialist

to associated Role Agents (RAs). Request for a set of Actor Agents must obey security restrictions given by the current step in the IIP and the current patient. Based on returned responses the RAs can offer to the PAs a best fit, or a sorted list of possible candidates for a human expert selection.

5. Further Improvements and Discussion

In the previous section we described the integration of the ProMA architecture into the K4Care system, which required only minimal conceptual changes in the model. The key advantage of this integration is in opening new possibilities for the K4Care system easily realized using ProMA. These, however, would require additional knowledge about the activities supported by the system.

The key factor is to use advanced simulation during the process of creating a new IIP. As we showed in the previous section, even a basic simulation can find some constraints that can be visualized and ease the process of creating a new IIP for a human expert.

Therefore, let us assume additional information available in the system. First of all, we can enhance the ontology for each action with additional properties, such as duration and type of the progress function (see [20] for more details on the progress function). Secondly, the map of a local area that the system is used in can be integrated into the multi-agent system as a virtual environment. Finally, we assume that the agent can access other relevant patients' data from the EHCR database during the process of creating the IIP (while obeying the security restrictions). All of these additional parameters allow us to simulate modelled procedures on a more detailed level. Based on a proposition of a new IIP a simulation can be ran and the results of the simulation can be used as an evaluation of the current version of the IIP and association of the actors.

The simulation of the IIP advance as in ProMA, which means that each of Process Agents is trying to successfully simulate associated step by selecting appropriate Role and Actor Agents.

As the Actors Agents are limited by other tasks already set in the system (i.e. there are several IIPs that are simulating at one time), they have to select the actions in which they currently participate according to their capabilities and priority. Let us demonstrate it on an example: we need to associate a nurse with an activity N.15 Blood sugar measurement at patient's home. We have a nurse Alice and a head nurse Eve. Nurse Alice has already planned a single task on the other side of the town, head nurse Eve has planned two shorter tasks in the centre. Note, that we use actors with different roles, however, there is hierarchy allowed within these roles and activity (i.e. a nurse is a more general role than the head nurse role). The ProMA architecture is able to handle this hierarchy and the RA correspondent to the nurse role would also send the proposition to RA correspondent to the head nurse role. In order to select an appropriate actor for this activity we use a multi-agent simulation in the virtual environment. That means the Actor Agents simulate execution of their already planned tasks within this environment. When the request for the N.15 activity is proposed by the RA associated with the nurse role, both Actor Agents are in different locations in the virtual environment and based on their other tasks they can estimate the time of completion of proposed task. Hence, in this simplified case, Eve would be offered for a human expert as an appropriate actor for this activity. However, if the proposed activity would have a higher priority for Alice, and her current task could be temporarily suspended, she could have been selected. In order to be the simulation of the steps as accurate as possible, PAs must have access to patients' health records to correctly estimate results of simulated activities (similarly to the critiquing mode of ProMA functioning describe in [19]) which can have impact of future course of execution of the IIP. Moreover, after the IIP is created and stored in the system, the ProMA can support critiquing and verifying the execution of these procedures and alert the health care specialist in charge in case of inconsistencies (see [19]).

6. Conclusions

In this paper we analysed the applicability of a generic multi-agent architecture into an existing decision support system, identified possible changes in both models and highlighted immediate advantages as well as possible future directions that are open by this integration. More specifically, we described processes as an appropriate notation for formalising medical procedural knowledge and described a process-based multi-agent architecture (ProMA). Then we focused on the existing support system, K4Care, where a procedural knowledge is stored in a SDA* formalism and used by a multi-agent system in order to provide support for home care specialists. We merged these two architectures using the minor model and conceptual changes and identified improvement in the process of creating a new Individual Intervention Plan for a patient. Finally, we identified possibilities for new features that the integration of ProMA into the K4Care system can further bring.

Acknowledgements

Acknowledgements belong to doc. Ing. L. Lhotska, CSc. for granting the access to K4Care documentation. The work was supported by the grant SVV-2010-265513, and by the Czech Ministry of Education, Youth and Sports under Research Programme no. MSM6840770038.

References

- [1] Balas E.A., Weingarten S., Garb C.T., Blumenthal D., Boren S.A., Brown G.D.: Improving preventive care by prompting physicians. *Arch Intern Med.* 2000 Feb 14;160(3):301308.
- [2] Garg A.X., Adhikari N.K., McDonald H. et al.: Effects of computerized clinical decision support systems on practitioner performance and patient outcomes: a systematic review. *JAMA.* 2005 Mar 9;293(10):12231238.
- [3] Grimshaw J.M., Russell I.T.: Effect of clinical guidelines on medical practice: a systematic review of rigorous evaluations. *Lancet.* 1993 Nov 27;342(8883):13171322.
- [4] Hunt D.L., Haynes R.B., Hanna S.E., Smith K.: Effects of computer-based clinical decision support systems on physician performance and patient outcomes: a systematic review. *JAMA.* 1998 Oct 21;280(15):13391346.
- [5] Kawamoto K., Houlihan C.A., Balas E.A., Lobach D.F.: Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success. *Bmj.* 2005 Apr 2;330(7494):765.
- [6] Burack R.C., Gimotty P.A.: Promoting screening mammography in inner-city settings: the sustained effectiveness of computerized reminders in a randomized controlled trial. *Med Care.* 1997;35:921-931.
- [7] Demakis J.G., Beauchamp C., Cull W.L. et al.: Improving residents' compliance with standards of ambulatory care: results from the VA Cooperative Study on Computerized Reminders. *JAMA.* 2000;284:1411-1416.
- [8] Overhage J.M., Tierney W.M., McDonald C.J.: Computer reminders to implement preventive care guidelines for hospitalized patients. *Arch Intern Med.* 1996;156:1551-1556.
- [9] Nilasena D.S., Lincoln M.J.: A computer-generated reminder system improves physician compliance with diabetes preventive care guidelines. *Proc Annu Symp Comput Appl Med Care.*
- [10] Egan M., Wells J., Byrne K., Jaglal S., Stolee P., Chesworth B. et al.: The process of decision-making in home-care case management: implications for the introduction of universal assessment and information technology. *Health & Social Care in the Community.* (2009, July); 17(4): 371-378.
- [11] Campana F., Moreno A., Riano D., Varga L.Z.: K4care: Knowledge-based homecare e-services for an ageing europe. *Agent Technology and e-Health.* 2008;p. 95115.
- [12] Aubrecht P., Lhotska L., Dolezel J., Dolezel J.: Mobile Devices for e-Services in Home Care. In: 4th European Conference of the International Federation for Medical and Biological Engineering. Springer; 2009. p. 10061009.
- [13] Isern D., Moreno A., Sanchez D., Hajnal A., Pedone G., Varga L.: Agent-based execution of personalised home care treatments. *Applied Intelligence.* 2009;p. 126.
- [14] Isern D., Moreno A.: Computer-based execution of clinical guidelines: A review. *International Journal of Medical Informatics.* 2008;77(12):787 808.
- [15] Isern D., Sanchez D., Moreno A.: HeCaSe2: A Multi-agent Ontology-Driven a Guideline Enactment Engine. In: CEEMAS '07: Proceedings of the 5th international Central and Eastern European conference on Multi-Agent Systems and Applications V. Berlin, Heidelberg: Springer-Verlag; 2007. p. 322324.
- [16] Novak P.: Jazzyk: A programming language for hybrid agents with heterogeneous knowledge representations. *Programming Multi-Agent Systems.* 2009;p. 7287.

- [17] Bordini R.H., Hubner J.F., Wooldridge M.J.: Programming multi-agent systems in AgentSpeak using Jason. Wiley-Interscience; 2007.
- [18] Hindriks K.V., De Boer F.S., Van der Hoek W., Meyer J.J.C.: Agent programming in 3APL. Autonomous Agents and Multi-Agent Systems. 1999;2(4):357401.
- [19] Bosansky B., Lhotska L.: Agent-based process-critiquing decision support system. In: 2nd International Symposium on Applied Sciences in Biomedical and Communication Technologies, 2009. ISABEL 2009; 2009. p. 16.
- [20] Bosansky B., Brom C.: Agent-Based Simulation of Business Processes in a Virtual World. In: Corchado E, Abraham A, Pedrycz W, editors. Hybrid Artificial Intelligence Systems. vol. 5271 of Lecture Notes in Computer Science. Springer Berlin / Heidelberg; 2008. p. 8694.
- [21] Riano D.: The SDA* Model: A Set Theory Approach. CBMS 2007, Maribor, Slovenia, (2007).
- [22] Lenz R., Blaser R., Beyer M., Heger O., Biber C., Baumlein M. et al.: IT support for clinical pathways Lessons learned. International Journal of Medical Informatics. 2007;76(Supplement 3):S397 S402.
- [23] Kumar A., Smith B., Pisanelli M., Gangemi A., Stefanelli M.: Clinical Guidelines as Plans: An Ontological Theory. Methods of Information in Medicine. 2006;2.
- [24] Fox J., Johns N., Rahmzadeh A., Thomson R. PROforma: A method and language for specifying clinical guidelines and protocols. In: Amsterdam; 1996.
- [25] Shahar Y., Miksch S., Johnson P.: The Asgaard project: a task-specific framework for the application and critiquing of time-oriented clinical guidelines. Artificial Intelligence in Medicine. 1998;14:2951.
- [26] Peleg M., Boxwala A., Ogunyemi O.: GLIF3: The Evolution of a Guideline Representation Format. Proc AMIA Annu Fall Symp. 2000;p. 645649.

Contact

Mgr. Branislav Bošanský

Agent Technology Center

Dept. Of Cybernetics

Faculty of Electrical Engineering

Czech Technical University in Prague

Technická 2

166 27 Prague 6

Czech Republic

e-mail: bosansky@labe.felk.cvut.cz

Temporomandibular Joint Prosthesis - 3D-CT Reconstruction before and after Treatment

E. Feltlová¹, T. Dostálová¹, M. Kašparová¹, J. Daněk², P. Hlíňáková¹, M. Hubáček¹, J. Nedoma³

¹Department of Paediatric Stomatology, Second Faculty of Medicine of Charles University, Prague, Czech Republic

²Department of Mathematics, University of West Bohemia, Plzen, Czech Republic

³Institute of Computer Science, Academy of Sciences of the Czech Republic, Prague, Czech

Supervisor: Prof. MUDr. Tatjana Dostálová, DrSc., MBA

Summary

The recent progress in understanding the biomedical basis for temporomandibular joint (TMJ) therapy as well as in computer and imaging technologies is beginning to provide novel insights into TMJ reconstruction, rational diagnostic and therapeutic strategies. Several methods are available for imaging the TMJ. These include basic radiography (such as panorex and corrected tomograms), ultrasonography, magnetic resonance imaging, and spiral or conebeam computed tomography. The latter technologies allow the joint to be visualized as sections in different planes and can also be rendered as 3D volumetric reconstructions to enhance diagnostic capabilities. The aim of the study is demonstrate patient's therapy including 1 year recall.

This contribution deals with 3D analysis of the mathematical and computer modeling of biomechanics of the jaw. Case report monitors step by step treatment plan, preprosthetic surgery, mandibular reconstruction and process of healing. 3D-CT was used to monitor the patient with dental cysts.

The mathematical and computer TMJ models analysis using overlapping of 3D CT before and after 1 year therapy shows the stability of TMJ partial prosthesis reconstruction.

Mathematical modeling and mathematical simulation of the reconstruction of patient jaw and/or patient TMJ may predict the joint loading of the reconstructed condyle. Our study demonstrated practical use of 3D images during TMJ therapy. Although many of the disorders involving the temporomandibular joint and associated structures can be diagnosed clinically on the basis of the history and physical findings, there are others that require the use of various imaging techniques to make

an accurate diagnosis or to determine the extent of involvement. Imaging the hard and soft tissues of the temporomandibular joint can be used not only for treatment planning but also long-term treatment evaluation.

Keywords: dentistry, maxillofacial surgery, partial TMJ prosthesis, three-dimensional computed tomography, magnetic resonance imaging, temporomandibular joint

1. Introduction

The temporomandibular joint (TMJ) is a bilateral composed joint connecting mandible to the temporal bone. Both joints function bilaterally and any change of one side influences the contra lateral joint. The TMJ has two articulating bone components - mandibular condyle and glenoid fossa of the temporal bone. The condyle has an ellipsoid shape with a mediolateral diameter of 20mm and anteroposterior diameter of 10 mm. The long axis of both condyles forms an angle 150-160°. The articular surface is covered with the fibrocartilage, instead of the hyaline cartilage usually present in synovial joints. The fibrocartilage is more resistant to degenerative changes and has better regeneration qualities. Another specialty of TMJ is its disc (discus-articularis) of a biconcave shape which fills the space between articular surfaces thus compensating their incongruities. The disc has a unique and complicated composition and it functions as a shock absorber. TMJs function symmetrically and this harmony allows biting, chewing and speaking. Movements are mostly combined together, resulting in the following jaw movements: depression, elevation of mandible (the opening and closing movement), propulsion (the forward movement), and lateropulsion (side movement) [1].

Temporomandibular disorders (TMDs) are a term embracing a number of clinical problems that involve the masticatory musculature, the temporomandibular joint and associated structures, or both. These disorders are accompanied by pain in the masticatory muscles, in the TMJ, and in the associated hard and soft tissues. Other symptoms include limitation or deviation in the mandibular range of motion, TMJ sounds, and/or headaches and facial pain [2]. Symptoms of TMDs occur in approximately 6 to 12 percent of the adult population or approximately ten million individuals in the United States [3]. Many of these patients can be managed with non-surgical therapies, but some end stage TMJ patients require surgical TMJ repair or reconstruction. The patient can develop end-stage TMJ as a result of trauma, osteoarthritis, reactive arthritis, ankylosis, idiopathic condylar resorption, connective tissue/ autoimmune diseases (e.g. rheumatoid arthritis, psoriatic arthritis, lupus, scleroderma, Sjogren's syndrome, ankylosing spondylitis) or other TMJ pathologies [4]. The joint damage may be followed by insertion of alloplastic implants. Joint replacement is a surgical procedure in which the severely damaged part of the TMJ is removed and replaced by a prosthetic device. If either a condyle or a fossa component of the TMJ is replaced, the surgery is called the partial joint replacement. In total joint replacement, condyle and fossa are both replaced [5]. The recent progress in understanding the biomedical basis for the TMJ therapy as well as in computer and imaging technologies is beginning to provide novel insights into TMJ reconstruction, rational diagnostic and therapeutic strategies. Several methods are available for imaging the TMJ.

These include basic radiography (such as panorex and corrected tomograms), ultrasonography, magnetic resonance imaging, and spiral or conebeam computed tomography. The latter technologies allow the joint to be visualized as sections in different planes and can also be rendered as 3D volumetric reconstructions to enhance diagnostic capabilities [6]. The aim of study is demonstrate patient's therapy including 1 year recall.

2. Case report

75-year-old patient, suffering from hypercholesterolemia, hypertension and allergy to analgesic Dinyll, was sent to our department from the surgical department of Ceske Budejovice hospital with suspicion of an extensive cyst of the mandible on the right side. From 2002 to 2009 she underwent 3 re-treatments at a private clinic because of an intraoral fistula in the same localization. Among the most important symptoms, which positively responded to antibiotic therapy, were difficult mouth opening, swelling and contracture. Repeated surgeries led to hypoesthesia of the right half of the lower lip. Unsuccessful treatment conducted to CT scan which revealed a large destructive cyst encroaching on the right mandibular ramus from the condyle to the angle of the mandible. During the first ambulatory check up in our clinic visible face asymmetry in the subzygomatic area on the right was clearly visible, the head of the right TM joint was not palpable, mouth opening was not limited, and the oral cavity was without pathological findings. The upper dental arch was toothless and in the lower dental arch teeth 33, 43, 44 were present. Dental rehabilitation in this phase of treatment was done by removable prosthetic appliances. Hypoesthesia of mandibular nerve on the right remained. Due to the extensive destruction of the mandibular ramus and its significant weakness in craniocaudal as well as mediolateral direction, which meant significant risk of pathological fractures, subtotal replacement of right TMJ was proposed (Fig. 1, 2, 3).

On June 22, 2009 under general anesthesia extirpation of the cyst of mandible on the right, partial resection of

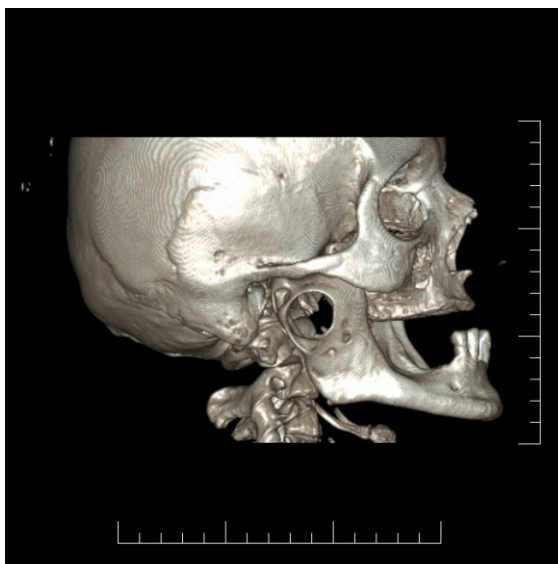


Fig. 1. 3D-CT analysis with destructive cyst in the right ramus mandibulae.

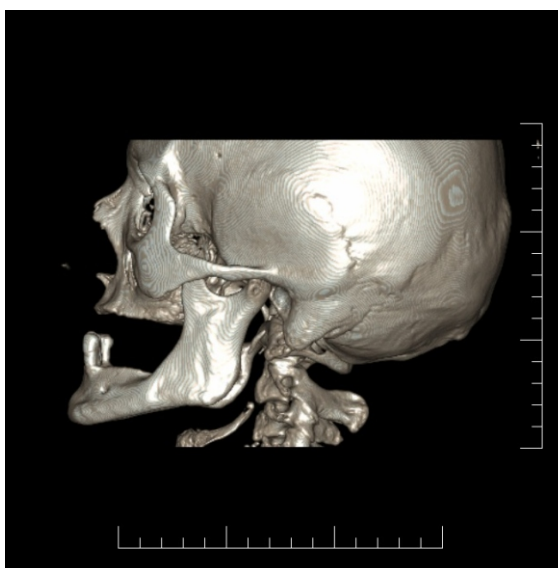


Fig. 2. 3D-CT analysis left healthy TMJ and ramus mandibulae.



Fig. 3. 3D-CT measurement of cast before therapy.

the body of mandible and exarticulation of the right TMJ followed by reconstruction of the right TMJ with subtotal replacement (UNI-LOCK) was made. The result of histological examination of biological material taken during the surgery confirmed the diagnosis of a radicular cyst. The postoperative period was covered by antibiotics for a total of 10 days. On the fourth day after the surgery the patient begins to open her mouth, and rehabilitation using wooden spatulas helped her to improve mouth opening. One week after the surgery she was discharged (Fig. 4). After discharge she was followed up in 3-week intervals. The patient was in a good shape, with no swelling, the wounds had healed, her mouth opening gradually improved from 15 mm to 30mm, the lower jaw did not turn aside when opening the mouth, palpation of the right TMJ was painless, however mild parenthesis of the mandible nerve persisted. On February 23, 2010 controlled CT scans were made. Results of the last CT scans were encouraging.

Thanks to the 3D-CT visualization of the case, the Surgery was reduced to a minimum, and after 1 year the patient reported minimal discomfort (Fig. 5, 6).

3. Methods

The patient data was processed (filtering, cropping, merging, iso-surface extraction, registration, etc.) and visualized using the proprietary software CTVolRegApp, that was developed at the University of West Bohemia under the Multimod Application Framework (MAF) [8], which is a visualisation system based mainly on VTK [7] but also on other specialised libraries. 3D-CT scanner will be used to monitor our patient with dental anomalies. 3D-CT scanners are based on volumetric tomography, using a 2D extended digital array providing an area detector. This is combined with a 3D X-ray beam. The cone-beam technique involves a single 360° scan in which the x-ray source and a reciprocating area detector synchronously move around the patient's head, which is stabilized with a head holder. At certain degree intervals, single projection images, known as "basis" images, are acquired.



Fig. 4. Partial TMJ prosthesis.



Fig. 5. 3D-CT partial TMJ prosthesis 1 year post insertion.



Fig. 6. 3D-CT left healthy TMJ and ramus mandibulae 1 year post insertion.

These are similar to lateral cephalometric radiographic images, each slightly offset from one another. This series of basis projection images is referred to as the projection data. Software programs incorporating sophisticated algorithms including back-filtered projection are applied to these image data to generate a 3D volumetric data set, which can be used to provide primary reconstruction images in 3 orthogonal planes (axial, sagittal and coronal). Although the 3D-CT principle has been in use for almost 2 decades, however, it is only recently, with the development of inexpensive x-ray tubes, high-quality detector systems and powerful personal computers, that affordable systems become commercially available [9].

4. Results

Data was organized regarding the presence or absence of bone changes in each view by each observer to allow precise comparison (Fig. 7-10). Fig. 7 shows the damaged right side of the joint, where the large radicular cyst was present. In the area of mandibular ramus and temporomandibular joint destruction of the bone is obvious and was replaced by subtotal prosthesis. Fig. 8 shows the healthy part of mandible. Fig. 9 shows the model of subtotal prosthesis and its fixation in the healthy bone tissue. The prosthesis is of the same anatomical shape as the joint. Two medical data sets of TMJ, before and after the treatment, were available for data processing. Each was obtained from various 3D-CT scanners and with a different quality. The position of the head during both CT scans was also different. For comparison of before and after the treatment we imported data sets into software programmers and generated the surface meshes for both models. Consequently the geometric transformations, i. e. translations and rotations, were used to find the best overlap of both geometries. The semitransparent colour was used to depict the original stage before the treatment for better imagination. Comparison of the jaw before and after surgery shows us changes in position in jaws in cranio-facial area.

5. Discussion

Mathematical and computer modeling of biomechanics of the jaw and the

temporomandibular joints have the potential for application in many areas of clinical practice, preprosthetic surgery and mandibular reconstruction. Clinical experiences have shown that patients who have had unilateral resection and subsequent reconstruction of the mandible have no pain associated with their

reconstructed condyle. Mathematical modeling and mathematical simulation of the reconstruction of patient jaw and/or patient TMJ may predict the joint loading of the reconstructed condyle. The other applications of mathematical modeling of jaw biomechanics can be useful with force changes and changes of their directions

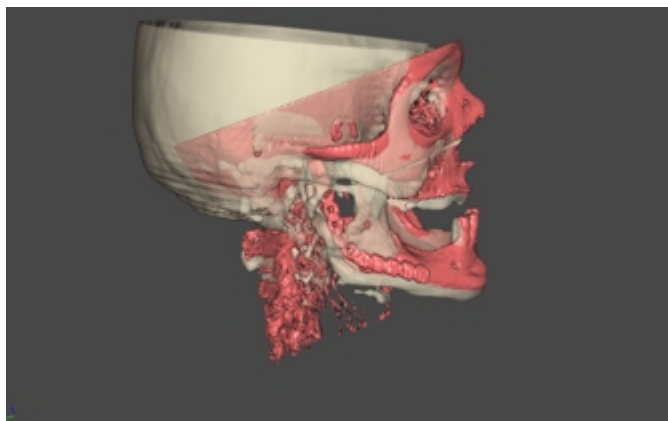


Fig. 7. 3D-CT comparison of the jaw before and after the surgery - right side.

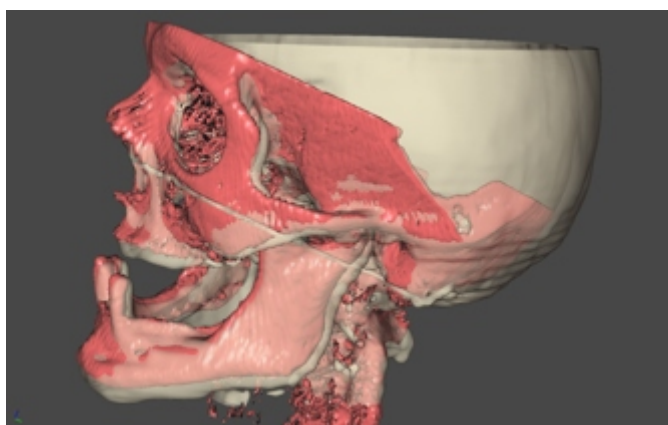


Fig. 8. 3D-CT comparison of the jaw before and after surgery - left side.

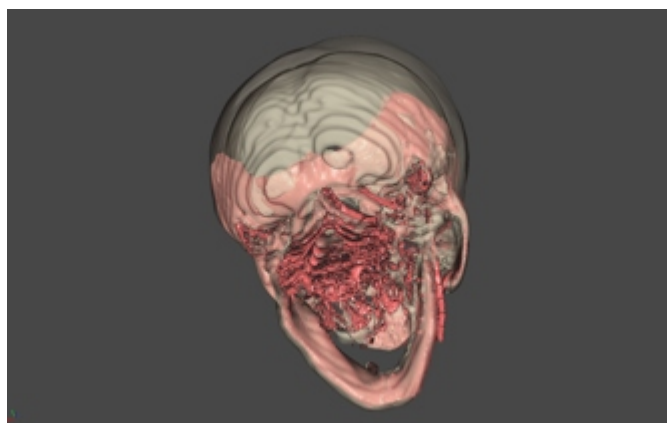


Fig. 9. 3D-CT comparison of the jaw before and after surgery - cranial sight.

after orthognathic surgery, and moreover, in evaluation of the influence of occlusal splints on TMJ loading as well as in prediction concerning with the tumor cancers or cysts. The main idea of our previous study continues in this project [10]. The development of CT techniques to acquire, process and manipulate images has improved the knowledge of structures that are otherwise difficult to visualize, such as the TMJ. Standardization of image formats using special protocols, and the use of free software follow a global trend. The manipulation of medical images should be mastered by health professionals who work with diagnosing, planning and treatment of patients, because this technology produces additional images to the ones initially reformatted by radiology technicians [11]. Our study demonstrated practical use of 3D images during the TMJ therapy. Although many of the disorders involving the temporomandibular joint and associated structures can be diagnosed clinically on the basis of the history and physical findings, there are others that require the use of various imaging techniques to make an accurate diagnosis or to determine the extent of involvement. Imaging the hard and soft tissues of the temporomandibular joint can be used not only for treatment planning but also long-term treatment evaluation.

Acknowledgments

This work was supported by the Grant 9991-3 IGA MZČR, FI-IM5/142 and SVV-2010-265513 projects.

References

- [1] Hlinakova P., Dostálová T., Danek J., Nedoma J.: „Temporomandibular joint and its two-dimensional and threedimensional modelling". *Mathematics and computers in simulation* Vol.80, pp. 1256- 2010, 2010.
- [2] Fikáčková H., Dostálová T., Navrátil L., Klaschka J.: Effectiveness of low-level laser therapy in temporomandibular joint disorders: A placebo-controlled study. *Photomedicine and Laser Surgery*, vol. 25, pp. 297-303, 2007.
- [3] Lipton J.A., Ship J.A., Larach-Robinson D.: Estimated prevalence and distribution of reported orofacial pain in the United States. *J Am Dent Assoc* vol. 124, pp. 11521, 1993.
- [4] Plesh O., Sinisi S.E., Crawford P. B., Gansky S.A.: Diagnoses based on the research diagnostic criteria for temporomandibular disorders in a biracial population of young women. *J Orofac Pain*. vol. 19, pp. 6575, 2005.
- [5] Wolford L.M., Karras S.C.: Autologous fat transplantation around temporomandibular joint total joint prostheses: preliminary treatment outcomes. *J. Oral Maxillofac Surg*. vol. 55, pp. 45251, 1997.
- [6] Wadhwa S., Kapila S.: TMJ disorders: future innovations in diagnostics and therapeutics. *J Dent Educ*. vol. 72, pp. 930947, 2008.
- [7] Viceconti M., Zannoni C., Testi D., Petrone M., Perticoni S., Quadrani P., Taddei F., Imboden S., Clapworthy G.J.: The Multimod Application Framework: a rapid application development tool for computer aided medicine. *Comput Methods Programs Biomed*, 85: 138-151, 2007
- [8] Schroeder W., Martin K., Lorensen B.: *The Visualization Toolkit: an object-oriented approach to 3D graphics*. Prentice Hall (1198)
- [9] Scarfe W. C.: Clinical applications of cone-beam computed tomography in dental practice. *J Can Dent Assoc*, vol. 72(1), pp. 7580, 2006.
- [10] Danek J., Hlinakova P., Preckova P., Dostalova T., Nedoma J.: Three-dimensional modelling of the temporomandibular joints and of their artificial replacements. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 6019 LNCS (PART 4), pp. 62-71, 2010
- [11] Farman A.G.: Raising standards: digital interoperability and DICOM. *Oral Surg Oral Med Oral Pathol Oral Radiol Endod*. vol. 99, pp. 525-6, 2005.

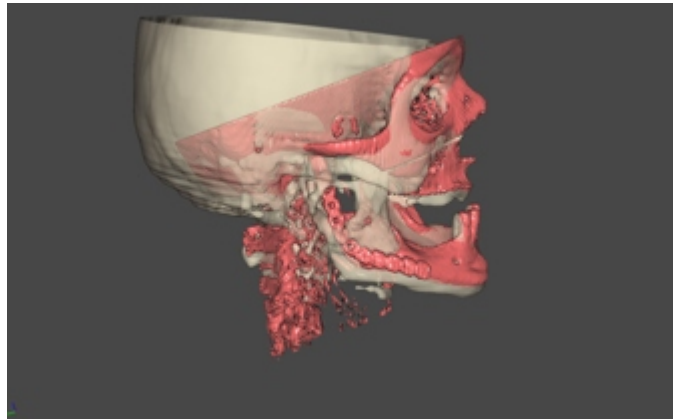


Fig. 10. 3D-CT cross-section after therapy.

Contact

MUDr. Eliška Feltlová

Department of Paediatric Stomatology
Second Faculty of Medicine of Charles University
V úvalu 84
15006 Prague 5
Czech Republic
e-mail: nelis.elis@seznam.cz

Methods of the Survival Analysis

J. Fürstová¹

¹Faculty of Medicine and Dentistry, Palacký University, Olomouc, Czech Republic

Supervisor: Doc. Zdeněk Valenta, M.Sc., M.S., Ph.D.

Summary

The survival analysis is a set of statistical methods dealing with time-to-event data. In biomedical applications the event of interest is usually relapse of the disease or death. A special feature of the survival analysis is censoring and truncation of data. When censoring or truncation occurs some information about the patients' survival is lost, e.g. some patients are lost to follow-up or the study ends before all the patients die. The survival analysis methods are used for estimation of the survival time distribution, for identification of risk factors that affect the survival time, and also for predicting the survival time when risk factors are present. Survival analysis methods have been further developed by the means of counting processes and martingale theory. Univariate survival analysis methods have been extended to multivariate setting. The multivariate survival analysis covers the field where independence between survival times cannot be assumed. Multi-state models and frailty models represent the two main approaches of multivariate methods.

Keywords: survival analysis, survival function, hazard function, cumulative hazard function, censoring, truncation, Kaplan-Meier estimator, Nelson-Aalen estimator, Cox PH model, partial likelihood, counting process, history, filtration, martingale, competing risk, multi-state models, frailty models

1. Introduction

The survival analysis is a collection of statistical methods for analyzing time-to-event data. The commencement of the survival analysis dates back to the 18th century when analyses of mortality experience of human populations started. During the World War II, the survival analysis focused on engineering - reliability of military equipment was being analyzed. After the World War II the interest turned towards economics and medicine. In 1960s, after the fundamental

article of E. L. Kaplan and P. Meier [9] had been published, medical applications of the survival analysis shifted to the center of statistical focus.

2. Basic concepts in the survival analysis

2.1 The survival and hazard function

Let X be the time until some specified event occurs, i.e. X is a non-negative real valued random variable having continuous distribution with finite expectation. There are several functions characterizing the distribution of X :

- The probability density of X : $f(x)$, $x \geq 0$.
- The survival function:

$$S(x) = P(X > x) = \int_x^\infty f(u) du \\ = 1 - F(x),$$

where $F(x)$ is the cumulative distribution function. The survival function describes the probability of an individual surviving beyond time x (experiencing the event after time x).

- The hazard function:

$$\lambda(x) = \lim_{\Delta x \rightarrow 0^+} \frac{P(x \leq X < x + \Delta x \mid X \geq x)}{\Delta x},$$

for all $x > 0$. The hazard function represents a conditional probability rate at which an individual alive at time x will experience an event in the next instant. There is a close relationship between the hazard and the survival functions:

$$\lambda(x) = \lim_{\Delta x \rightarrow 0^+} \frac{1}{\Delta x} \frac{P(x \leq X < x + \Delta x)}{P(X \geq x)} \\ = \lim_{\Delta x \rightarrow 0^+} \frac{\frac{1}{\Delta x} \int_x^{x+\Delta x} f(u) du}{\int_x^\infty f(u) du} = \frac{f(x)}{S(x)}$$

$$\frac{dS(x)}{dx} = -\frac{f(x)}{S(x)} = -\frac{d}{dx} \ln S(x).$$

- The cumulative hazard function:

$$\Lambda(x) = \int_0^x \lambda(u) du = -\ln S(x).$$

Thus

$$S(x) = \exp(-\Lambda(x)) = \exp\left(-\int_0^x \lambda(u) du\right).$$

If X is a discrete random variable taking values $x_1 < x_2 < \dots$ with associated probability mass function $f(x_i) = P(X = x_i)$, $i = 1, 2, \dots$, the survival function is

$$S(x) = \sum_{j: x_j > x} f(x_j)$$

the hazard at x_i is

$$\lambda_i = P(X = x_i \mid X \geq x_i) = \frac{f(x_i)}{S(x_i^-)}, \quad i = 1, 2, \dots,$$

where $S(x^-) = \lim_{t \rightarrow x^-} S(t)$. The survival function and probability mass function can be also written as (see [8])

$$S(x) = \prod_{j: x_j \leq x} (1 - \lambda_j), \quad f(x_i) = \lambda_i \prod_{j=1}^{i-1} (1 - \lambda_j).$$

More generally, the distribution of X may have both discrete and continuous components. The approach to the discrete and continuous parts can be unified through the notion of a product integral: Let $\lambda_c(x)$ be the continuous component of the hazard function, and let $\lambda_1, \lambda_2, \dots$ be the discrete components at times $x_1 < x_2 < \dots$. The overall survival function is then

$$S(x) = \exp\left(-\int_0^x \lambda_c(u) du\right) \prod_{j: x_j \leq x} (1 - \lambda_j)$$

and the cumulative hazard function is

$$\Lambda(x) = \int_0^x \lambda_c(u) du + \sum_{j: x_j \leq x} \ln(1 - \lambda_j).$$

Let $d\Lambda(x)$ be a differential increment of the cumulative hazard over the interval $[x, x + dx)$:

$$\begin{aligned} d\Lambda(x) &= \Lambda(x^- + dx) - \Lambda(x^-) \\ &= P(X \in [x, x + dx) | X \geq x) \\ &= \begin{cases} -\ln(1 - \lambda_i) & \text{for } x = x_i, i = 1, 2, \dots \\ \lambda_c(x)dx & \text{otherwise.} \end{cases} \end{aligned}$$

The survival function in the discrete, continuous, or mixed cases can then be written as

$$S(x) = P_0^x(1 - d\Lambda(u)),$$

where

$$P_0^x(1 - d\Lambda(u)) = \lim_{r \rightarrow \infty} \prod_{k=1}^r (1 - (\Lambda(u_k) - \Lambda(u_{k-1})))$$

is the product integral [8].

2.2 Censoring and truncation

Survival data possess a special feature of censoring, compared to other statistical data. Censoring is used when the survival time is not known exactly, the event is only known to have occurred within some time interval. There are several types of censoring: right, left and interval. In biomedical applications, right censoring is the most common type of censoring. It occurs when the survival time is incomplete on the right-hand side of the follow-up period, i.e. the study ends before all patients experience the event or a patient is lost to follow-up (dies due to reasons other than the event of interest, withdraws from the study, moves to another city, etc.).

Let X_1, X_2, \dots, X_n be independent and identically distributed (i.i.d.) survival times and C_1, C_2, \dots, C_n be i.i.d. censoring times. The lifetime X_i of the i -th individual will be known if, and only if, $X_i < C_i$. If $C_i < X_i$ the event time will be censored at C_i . Thus it is convenient to represent the survival experience of a group of patients by the pairs of random variables (T_i, δ_i) where $T_i = \min(X_i, C_i)$, $\delta_i = I(X_i < C_i)$ and I is an indicator of the event's occurring, having value one if the event occurs, and zero otherwise.

Another feature, common in survival data, is a truncation. Truncation occurs when

only those individuals whose event time lies within a certain time interval (T_L, T_R) are observed. For left truncation, $T_R = \infty$ in case of right truncation, $T_L = 0$. Individuals, whose event time is not in this interval, are not observed and no information on these subjects is available. This is in contrast to censoring where there is at least partial information available on each patient. When data are truncated, a conditional distribution has to be used in constructing the likelihood (see [10]).

A critical assumption for the likelihood construction is the independence of lifetimes and censoring times. Censoring is said to be independent if the failure rates that apply to individuals on trial at each time $t > 0$ are the same as those that would have applied had there been no censoring [8]. Thus the requirement is that at each time t

$$\lim_{\Delta t \rightarrow 0} \frac{P(T \in [t, t + \Delta t) | T \geq t)}{\Delta t} = \lim_{\Delta t \rightarrow 0} \frac{P(T \in [t, t + \Delta t) | T \geq t, Y(t) = 1)}{\Delta t},$$

where $Y(t) = 1$ indicates that the individual is at risk of failure at time t (has neither failed nor been censored prior to t).

2.3 Counting processes and martingales

An alternative approach to develop inference procedures for censored data involves counting processes. A counting process $N = \{N(t), t \geq 0\}$ is a stochastic process with $N(0) = 0$ whose value at time t counts the number of events that have occurred in the interval $(0, t]$. The sample paths (realizations) of N are nondecreasing, right-continuous step functions that jump whenever an event (or events) occur. In the counting process formulation, the pair of variables (T_i, δ_i) introduced in Section 2.2 is replaced with the pair of functions $N_i(t), Y_i(t), i = 1, \dots, n$ where

$N_i(t)$ = no. of events observed in $[0, t]$ for unit i

$$Y_i(t) = \begin{cases} 1 & \text{unit } i \text{ is at risk at time } t, \\ 0 & \text{otherwise.} \end{cases}$$

$N_i(t)$ is a counting process, while $Y_i(t)$ is a predictable process, i.e. a process whose value at time t is known infinitesimally before t at time t . This process has left-continuous sample paths. Right-censored survival data are included in this formulation as a special case:

$$N_i(t) = I(T_i \leq t, \delta_i = 1) \text{ and } Y_i(t) = I(T_i \geq t).$$

To deal with all on-study information of each patient, a term history (or filtration) is used. A history, denoted $\{F_t, t \geq 0\}$ is a σ -algebra generated by N_i and Y_i :

$$F_t = \sigma(N_i(s), Y_i(s^+), i = 1, \dots, n; 0 \leq s \leq t),$$

where $Y_i(s^+) = \lim_{u \rightarrow s^+} Y_i(u)$.

Thus F_t contains the information up to and including time t . The information in F_t increases with increasing time on study, i.e. $F_s \subseteq F_t$ for $s \leq t$ [4]. Let $dN_i(t)$ denote the increment of over the time interval $[t, t + dt)$:

$$dN_i(t) = N_i((t + dt)^-) - N_i(t^-).$$

For each $t > 0$ let

$$F_{t-} = \sigma(N_i(s), Y_i(s), i = 1, \dots, n; 0 \leq s < t)$$

denote the full history of the processes $N_i(s), Y_i(s), i = 1, \dots, n$ up to but not including t . Then (see [4]):

$$E(dN_i(t) | F_{t-}) = Y_i(t)\lambda_i(t)dt,$$

where $\lambda_i(t)$ is the hazard function. The process

$$\Lambda_i(t) = \int_0^t Y_i(s)\lambda_i(s)ds, \quad t \geq 0,$$

is called the intensity process. At each fixed t , this process is a random variable which approximates the number of jumps by N_i over $(0, t]$. In fact, $EN_i(t) = E\Lambda_i(t)$ and thus [4].

$$E(N_i(t) | F_{t-}) = E(\Lambda_i(t) | F_{t-}) = \Lambda_i(t)$$

For any given i define the process

$$M_i(t) = N_i(t) - \int_0^t Y_i(s)\lambda_i(s)ds, \quad t \geq 0, i = 1, \dots, n. \quad (1)$$

Equivalently, the process can be defined

$$M_i(t) = \int_0^t dM_i(s),$$

where

$$dM_i(t) = dN_i(t) - Y_i(t)\lambda_i(t)dt.$$

It can be seen that $E(dM_i(t)|F_t) = 0$ for all t and $E(M_i(t)|F_s) = M_i(s)$ for all $s \leq t$ [4]. A process that satisfies these (equivalent) conditions is a martingale. According to the Doob-Meier decomposition theorem (see [4]), any counting process may be uniquely decomposed as a sum of a martingale and a compensator C which is a predictable, right-continuous process with $C(0) = 0$. As an example, according to (1)

$$\begin{aligned} N_i(t) &= M_i(t) + \int_0^t Y_i(s) \lambda_i(s) ds \\ &= M_i(t) + \Lambda_i(t), \end{aligned} \quad (2)$$

where $M_i(t)$ is the counting process martingale corresponding to $N_i(t)$ and $\Lambda_i(t)$ is the compensator of the counting process N_i with respect to the filtration F_t . In terms of differential increments, the process (2) can be equivalently written as

$$dN_i(t) = dM_i(t) + Y_i(t) \lambda_i(t) dt.$$

The approach using martingale methods is very useful in yielding results for censored and truncated data, especially for calculating and verifying asymptotic properties of test statistics and estimators.

3. Non-parametric and semi-parametric models

A principle objective of the survival analysis focuses on estimation of basic quantities (the survival and hazard function) based on censored data. To analyze survival data parametrically, assumptions about the distribution of the failure times would have to be made. To avoid such assumptions, it is common to use non-parametric models. The simplest non-parametric estimate of a distribution function is the empirical distribution function

$$F_n(x) = \frac{\text{no. of sample values} \leq x}{n},$$

when a continuous distribution is estimated by a discrete one. For an uncensored sample of n distinct failure times, the empirical survival function is then estimated by $S_n(t) = 1 - F_n(t)$. The only problem with this approach is the censoring - it is not taken into account in standard statistical methods. Important steps in the development of appropriate

methods were done by Kaplan and Meier [9] and Cox [3].

3.1 Kaplan-Meier and Nelson-Aalen estimators

The Kaplan-Meier estimator (called also the product-limit estimator) estimates the survival function by

$$\hat{S}(t) = \prod_{i: t_i \leq t} \left(1 - \frac{d_i}{R_i}\right),$$

where d_i there are events observed at time t_i and R_i is the number of individuals still at risk at time t_i (uncensored survivors just before t_i). The variance of the estimator can be estimated using Greenwood's formula (see [11]):

$$\hat{V}(\hat{S}(t)) = \hat{S}(t)^2 \sum_{i: t_i \leq t} \frac{d_i}{R_i(R_i - d_i)}.$$

The product $\hat{\Lambda}(t) = -\ln(\hat{S}(t))$ can also be used to estimate the cumulative hazard function:

An alternative estimator of the cumulative hazard function was proposed by Nelson in 1972 [12] and rediscovered by Aalen in 1978 [1]:

$$\tilde{\Lambda}(t) = \sum_{i: t_i \leq t} \frac{d_i}{R_i}.$$

The variance of the Nelson-Aalen estimator was estimated by Aalen using counting process techniques and is given by

$$\tilde{V}(\tilde{\Lambda}(t)) = \sum_{i: t_i \leq t} \frac{d_i}{R_i^2}.$$

Based on the Nelson-Aalen estimator of the cumulative hazard function, an alternative estimator of the survival function becomes

$$\tilde{S}(t) = \exp(-\tilde{\Lambda}(t)).$$

Suppose now that n individuals from a homogeneous population are put on a study at time 0. Let N_i be the counting process and Y_i be the at-risk process of the i -th individual, as described in Section 2.3. Let

$$N_*(t) = \sum_{i=1}^n N_i(t) \quad \text{and} \quad Y_*(t) = \sum_{i=1}^n Y_i(t), \quad 0 < t < \infty.$$

$N_*(t)$ denotes the total number of observed

failures in the interval, while $Y_*(t)$ is the number of individuals in the entire study group that are at risk at time t . The Nelson-Aalen estimator of the cumulative hazard can be written in the counting process notation as

$$\tilde{\Lambda}(t) = \int_0^t \frac{J(u)}{Y_*(u)} dN_*(u),$$

where $J(u) = I(Y_*(u) > 0)$ with the convention that $0/0$ is interpreted as 0 [8]. The Kaplan-Meier estimator of the survival function is then

$$\hat{S}(t) = \prod_{u \leq t} (1 - d\tilde{\Lambda}(u)).$$

A common interest is to compare two or more samples, i.e. to test whether there is a significant difference in survival experience of distinct groups of patients. Several generalizations of standard non-parametric tests have been developed to deal with censored and truncated data. The most common tests are the log-rank test, Gehan-Wilcoxon test and Peto-Peto test. For more information, see e.g. [10].

3.2 The Cox model

In clinical studies we typically examine the association of several risk factors with the occurrence of the event of interest. Various patients' characteristics may be associated with patients' survival experience (e.g. age, sex, blood pressure, ...). The Cox proportional hazards model has become a popular approach to modeling covariate effects on survival. In this model the intensity process (hazard) for the i -th subject is

$$\lambda_i(t) = Y_i(t) \lambda_0(t) \exp(\beta^T X_i),$$

where $Y_i(t)$ is the at-risk process, λ_0 is the baseline hazard (common to all individuals in the study population), X_i is the vector of covariates of individual i and β is a vector of unknown regression parameters. In this model, the ratio of hazard functions of two individuals is constant (the baseline hazard $\lambda_0(t)$ is canceled out), thus the temporal effect is separated from the effect of the covariates. Estimation of the regression coefficients is based on maximizing of the partial likelihood function, which was introduced by Cox in 1972 [3].

The partial likelihood function for β reads

$$L(\beta) = \prod_{j=1}^k \frac{\exp(\beta^T X_j)}{\sum_{l \in R(t_j)} \exp(\beta^T X_l)},$$

where $t_1 < \dots < t_k$ are the uncensored failure times of the study group, $R(t)$ is the set of subjects at risk of failure at time t (just prior to time t), and X_j denotes the covariate vector for an individual failing at t_j . The partial likelihood function is treated as a standard likelihood, and inference is carried out by usual means.

4. Multivariate survival analysis

In most clinical applications the univariate survival analysis assumes that the observed survival times are mutually independent (i.i.d. failure times). In practice, however, dependence can occur for very different kinds of data, e.g. survival of twins or other several individuals, similar organs, recurrent events or multi-state events. The multivariate survival analysis covers the field where independence between survival times cannot be assumed. According to [7], the various approaches to analyzing multivariate survival data fall into four main categories: multi-state models, frailty models, marginal modeling and non-parametric methods. The data structure should be considered as well. The data can be parallel (where the number of failures is fixed by the design of the study) or longitudinal (where the number of failures is random for each object under study). The data sets are classified into six types: several individuals, similar organs, recurrent events, repeated measurements, different events and competing risks. Relation of the data types to the two main approaches of analysis (multi-state and frailty models) is described in Table 1. Only these two approaches to analyzing multivariate survival data are presented in this paper. For more information on marginal and non-parametric methods, see e.g. [7], [16], or [8].

4.1 Competing risk and multi-state models

Multi-state models are commonly used for describing the development of longitudinal data. They model stochastic processes, which at any time point occupy one of a set

Tab. 1. Overview of data types and approaches; x means relevant, blank not relevant. Adopted from [7].

Type of data	Multi-state	Frailty
Several individuals	x	x
Similar organs	x	x
Recurrent events	x	x
Repeated measurements		x
Different events	x	
Competing risks	x	

of discrete states. In medicine, the states can be e.g. healthy, diseased, and dead. A change of state is called a transition. The competing risk model is an example of multi-state modeling. In competing risks, various causes of death “compete” in the life of patient, and occurrence of one event precludes occurrence of the other events. There are generally three areas of interest in the analysis of competing risks [8]:

1. Studying the relationship between a vector of covariates and the rate of occurrence of specific types of failure.
2. Analyzing whether patients at high risk of one type of failure are also at high risk for others.
3. Estimating the risk of one type of failure after removing others.

Suppose that individuals under study can experience any one of m distinct failure types. For each individual, the underlying failure time T and a covariate vector X are known. The overall hazard function at time t is

$$\lambda(t, X) = \lim_{\Delta t \rightarrow 0^+} \frac{P(t \leq T < t + \Delta t | T \geq t, X)}{\Delta t}.$$

To model competing risks, a cause-specific hazard function is considered:

$$\lambda_j(t, X) = \lim_{\Delta t \rightarrow 0^+} \frac{P(t \leq T < t + \Delta t, J = j | T \geq t, X)}{\Delta t}$$

for $j = 1, \dots, m$ is a random variable representing the type of failure, and $t > 0$. In words, $\lambda_j(t, X)$ specifies the rate of type j failures, given and in the presence of all other failure types [8]. If only one of the failure types can occur, then

$$\lambda(t, X) = \sum_{j=1}^m \lambda_j(t, X)$$

due to the law of total probability.

It is possible to calculate the Kaplan-Meier estimator for each type of failure separately, but it is difficult to give this a survival function interpretation and therefore this is not recommended [7]. Instead, generalizations of the Kaplan-Meier and Nelson-Aalen estimators can be made (see e.g. [8]). The generalized estimator includes all causes of failure and is usually denoted the Aalen-Johanson estimator.

The Cox model for the cause-specific hazard functions can be considered:

$$\lambda_j(t, X) = \lambda_{0j}(t) \exp(\beta_j^T X), \quad j = 1, \dots, m.$$

Both the baseline hazards λ_{0j} and the regression coefficients β_j vary arbitrarily over the failure types. Estimation and comparison of the coefficients β_j can be conducted by applying asymptotic likelihood techniques individually to the m factors.

A traditional approach to multi-state models is based on the Markov models. Consider first a homogeneous population with no covariates. Let $A(t)$ be the state occupied at time $t \geq 0$ with probability model of $A(t)$ being the Markov process. The individuals under study move among $m > 1$ discrete states.

If a randomly chosen individual is in state i at time t , the transition rate (or intensity) from i to j at time t is given by

$$d\Lambda_{ij}(t) = P[A(t^- + dt) = j \mid A(u), 0 \leq u < t, A(t^-) = i] \\ = P[A(t^- + dt) = j \mid A(t^-) = i], \quad t > 0,$$

which holds for all $A(u)$, $0 \leq u < t$ with $A(t) = i$ and $i, j \in \{1, \dots, m\}$, $j \neq i$. The process is memoryless in that only the current state occupied is relevant in specifying the transition rates [8]. In the continuous case, $d\lambda_{ij}(t) = \lambda_{ij}(t)dt$ for all $i, j = 1, \dots, m$ so that $\lambda_{ij}(t)$, $i \neq j$ is the continuous-time intensity function for i -to- j transitions. Estimation of the cumulative intensity functions $\lambda_{ij}(t)$ proceeds as follows [8]: consider a possibly right-censored sample of n individuals. For $k = 1, \dots, n$, let $N_{ijk}(t)$ be the right continuous process that counts the number of observed direct i -to- j transitions for k -th individual, $i, j = 1, \dots, m$, $i \neq j$. Let $Y_{ik}(t)$ be the corresponding at-risk process. Define the filtration process as

$$F_t = (N_{ijk}(t), Y_{ik}(u^+), 0 \leq u \leq t),$$

For $k = 1, \dots, n$; $i, j = 1, \dots, m$ and suppose that censoring is independent, so that

$$P(dN_{ijk}(t) = 1 \mid F_{t-}) = Y_{ik}(t)d\Lambda_{ij}(t),$$

which must hold for all i, j, k and $t > 0$. The Nelson-Aalen estimator of $\lambda_{ij}(t)$ is then given by

$$d\hat{\Lambda}_{ij}(t) = \frac{dN_{ij \cdot}(t)}{Y_{i \cdot}(t)}$$

for all $i \neq j$.

When the vector of covariates X is present, the continuous-time modulated Markov model can be specified for the underlying intensity function

$$\lambda_{ijk}(t) = \lim_{dt \rightarrow 0^+} \frac{P(A_k(t^- + dt) = j \mid A_k(t^-) = i, X)}{dt}.$$

Parametric and semi-parametric models for λ_{ijk} are obtained analogously as earlier and may be found in [8].

4.2 Frailty models

Frailty models represent an extension of the Cox proportional hazards model. The concept of frailty provides a way to introduce random effects into the model to account for association (correlation) and unobserved heterogeneity. This heterogeneity may be difficult to assess but is nevertheless of a great importance. The frailty is an unobserved random factor that modifies multiplicatively the hazard function of an individual or a group of individuals. The key idea of these models is that individuals most "frail" die earlier than the others [16]. The frailty models are relevant to lifetimes of several individuals, similar organs and repeated measurements. They are not generally relevant for the case of different events [7].

First, bivariate models will be considered. Let

$$S_{12}(t_1, t_2) = P(T_1 \geq t_1, T_2 \geq t_2)$$

be the joint survival function for the two survival times T_1 and T_2 where $S_{12}(t, t)$ is the probability that both subjects under study will be alive at time t .

The marginal survival functions are then

$$S_1(t_1) = P(T_1 \geq t_1) = S_{12}(t_1, 0)$$

$$S_2(t_2) = P(T_2 \geq t_2) = S_{12}(0, t_2).$$

If T_1 and T_2 are independent, $S_{12}(t_1, t_2) = S_1(t_1)S_2(t_2)$. The joint hazard function is

$$\lambda_{12}(t_1, t_2) = \lim_{\Delta t \rightarrow 0^+} \frac{P(T_1 \in [t_1, t_1 + \Delta t), T_2 \in [t_2, t_2 + \Delta t) \mid T_1 \geq t_1)}{\Delta t^2},$$

and the marginal hazards are

$$\lambda_i(t_i) = \lim_{\Delta t \rightarrow 0^+} \frac{P(T_i \in [t_i, t_i + \Delta t) \mid T_i \geq t_i)}{\Delta t},$$

For $i = 1, 2$. To address heterogeneity in the survival times it is assumed that the lifetimes are conditionally independent, i.e. T_1 and T_2 are independent given the random effect Z called frailty:

$$S_{12}(t_1, t_2 \mid Z) = S_1(t_1 \mid Z)S_2(t_2 \mid Z).$$

Usually, the frailty is assumed to act multiplicatively on the hazard, so that

$$\lambda_i(t_i) = Z\lambda_{0i}(t_i) \quad \text{and} \quad S_i(t_i \mid Z) = S_{0i}(t_i)^Z$$

for some baseline hazard $\lambda_{0i}(t)$ and baseline survival function $S_{0i}(t)$ (when known covariates X_i are present, the hazard may be expressed as

$$\lambda_{0i}(t_i) = \lambda_0(t_i) \exp(\beta^T X_i)$$

through the Cox regression model). Under the assumption of multiplicative frailty, the cumulative hazards are

$$\Lambda_i(t_i) = Z\Lambda_{0i}(t_i).$$

The conditional joint survival function is then

$$S_{12}(t_1, t_2 \mid Z) = S_{01}(t_1)^Z S_{02}(t_2)^Z \\ = \exp(-Z\Lambda_{01}(t_1)) \exp(-Z\Lambda_{02}(t_2)) \\ = \exp(-Z(\Lambda_{01}(t_1) + \Lambda_{02}(t_2))).$$

As the frailty Z is an unobserved effect, it needs to be 'integrated out' of the survival function. This is done by the Laplace transform, which is defined for a random variable Z as

$$L_g(s) = \int \exp(-sz)g(z)dz = E(\exp(-sZ)).$$

Where $g(z)$ is the probability density of Z . For the bivariate survival function thus

$$S_{12}(t_1, t_2) = \int_0^\infty S_{12}(t_1, t_2 \mid Z)g(z)dz \\ = \int_0^\infty \exp(-Z(\Lambda_{01}(t_1) + \Lambda_{02}(t_2)))g(z)dz \\ = L_g(\Lambda_{01}(t_1) + \Lambda_{02}(t_2)),$$

where the Laplace transform of $g(z)$ is evaluated at

$$s = \Lambda_{01}(t_1) + \Lambda_{02}(t_2).$$

In many applications, the frailty Z is assumed to follow some distribution with the explicit Laplace transform. A standard (and most widely used) distribution for frailty is the gamma distribution.

The random variable Z is gamma distributed with parameters k and Θ ($Z \sim \Gamma(k, \Theta)$), if its probability density function is

$$g(z) = \frac{\theta^k z^{k-1} \exp(-\theta z)}{\Gamma(k)}, \quad k, \theta > 0 \text{ and } z > 0,$$

$$\text{With } EZ = \frac{k}{\theta}, \quad \text{var}Z = \frac{k}{\theta^2}.$$

The gamma function in the denominator of the probability density function is defined as

$$\Gamma(k) = \int_0^\infty u^{k-1} \exp(-u) du, \quad \text{for } k > 0.$$

It satisfies $\Gamma(k+1) = k\Gamma(k)$.

The gamma distribution fits very well to failure data and is also convenient from computational and analytical point of views [19].

Suppose the common frailty component Z has a gamma distribution with parameters $k = \Theta = 1/\sigma^2$. The Laplace transform of the gamma density is then

$$L(s) = \left(\frac{\theta}{\theta + s} \right)^k,$$

Which leads to (see [7])

$$S_{12}(t_1, t_2) = \left(\frac{1}{1 + \sigma^2 \Lambda_{01}(t_1) + \sigma^2 \Lambda_{02}(t_2)} \right)^{1/\sigma^2}.$$

To extend the bivariate model to a multivariate one, consider a set of clustered data where for the j -th individual in the i -th group (or cluster) there are the observation times t_{ij} and the vector of covariates X_{ij} . The assumption is, again, that given X_{ij} and a random effect Z_i the m_i lifetimes in group i are independent. Thus the joint distribution of these lifetimes given Z_i is the product of the marginal distributions given Z_i . The marginal hazards then satisfy

$$\lambda_{ij}(t_{ij} | X_{ij}, Z_i) = Z_i \lambda_{0ij}(t_{ij} | X_{ij}).$$

When the hazards are modeled using the Cox proportional hazards,

$$\lambda_{0ij}(t_{ij} | X_{ij}) = \lambda_0(t_{ij}) \exp(\beta^T X_{ij}).$$

If the cluster-specific random effects Z_i have independent gamma distributions, then the unconditional survival for the m_i lifetimes in cluster is

$$S_i(t_i, X_i) = \int_0^\infty \prod_j S_{ij}(t_{ij} | X_{ij}, Z_i) g(z_i) dz_i,$$

where $t_i = (t_{i1}, t_{i2}, \dots, t_{im_i})^T$, $X_i = (X_{ij})_{m_i \times n}$

This can be solved using the Laplace transform (see [14])

$$S(t_i, X_i) = \left(\frac{1}{\Psi} \right)^{1/\sigma^2},$$

Where

$$\Psi = 1 + \sigma^2 \Lambda_0(t_{i1}) \exp(\beta^T X_{i1}) + \dots + \sigma^2 \Lambda_0(t_{im_i}) \exp(\beta^T X_{im_i}).$$

Different choices of distribution for the frailty Z are possible, e.g. the family of positive stable distributions or the PVF (power variance function) family. For more information about these, see [7]. The frailty Z may also be treated non-parametrically. Although it is desirable to have completely non-parametric estimate of the survival function, the estimates are mathematically complicated and are not of major importance [7].

Statistical models that use counting process notation and are convenient for these types of analyses are slightly different from those used until now. In the previously used models, the intensity process $\lambda(t)$ at the follow-up time t given the covariates X was

$$\lambda(t) dt = P(dN(t) = 1 | N(s), 0 \leq s < t, X).$$

In this expression it is assumed that jumps in N are of a unit size only. However, recurrent and correlated failure time data include jumps of a size greater than one (more than one event can be recorded for an individual at a specific follow-up time). Thus it is natural to model the mean jump in N across time:

$$d\Lambda(t) = E(dN(t) | N(s), 0 \leq s < t, X).$$

in the cumulative intensity process. The Cox-type model for the intensity process is then

$$d\Lambda(t) = d\Lambda_0(t) \exp(\beta^T X).$$

For more details, see [8].

5. Conclusion

The survival analysis is a collection of specific statistical methods. In this paper, a short overview of these methods was presented. The standard univariate models were extended to multivariate models dealing with parallel and longitudinal data. The two major multivariate concepts were introduced: multi-state and frailty models.

Acknowledgments

The paper has been supported by the project SVV-2010-265513.

References

- [1] Aalen O. O.: Nonparametric Inference for a Family of Counting Processes. *Annals of Statistics* 6 (1978), 701726.
- [2] Andersen P. K., Gill R. D.: Cox's regression model for counting processes: a large sample study. *The Annals of Statistics* 10 (1982), 11001120.
- [3] Cox D. R.: Regression Models and Life-Tables. *Journal of the Royal Statistical Society B* 34 (1972), 187220.
- [4] Fleming T. R., Harrington D. P.: Counting Processes and Survival Analysis. John Wiley & Sons, New York, 1991.
- [5] Fúrstová J.: Multivariate Methods of Survival Analysis. Doktorandský den 2010, Matfyzpress, Praha, 2010.
- [6] Gill R. D.: Understanding Cox's regression model: a martingale approach. *Journal of the American Statistical Association* 79 (1984), 441447.
- [7] Hougaard P.: Analysis of Multivariate Survival Data. Springer, New York, 2000.
- [8] Kalbfleisch J. D., Prentice R. L.: The Statistical Analysis of Failure Time Data. John Wiley & Sons, New York, 2002.
- [9] Kaplan E. L., Meier P.: Nonparametric Estimation from Incomplete Observations. *Journal of the American Statistical Association* 53 (1958), 457481.
- [10] Klein J. P., Moeschberger M. L.: Survival Analysis. Techniques for Censored and Truncated Data. Springer, New York, 2003.
- [11] Miller R. G., Gong G., Muñoz A.: Survival Analysis. John Wiley & Sons, New York, 1998.
- [12] Nelson W.: Theory and Applications of Hazard Plotting for Censored Failure Data. *Technometrics* 14 (1972), 945965.
- [13] Prentice R. L., Williams B. J., Peterson A. V.: On the regression analysis of multivariate failure time data. *Biometrika* 68 (1981), 373379.

- [14] Rodríguez G.: Multivariate Survival Models. available at <http://data.princeton.edu/>, cited on April 10, 2010.
- [15] Self S. G., Prentice R. L.: Commentary on Andersen and Gill's "Cox's regression model for counting processes: a large sample study". The Annals of Statistics 10 (1982), 11211124.
- [16] Therneau T. M., Grambsch P. M.: Modeling Survival Data. Extending the Cox Model. Springer, New York, 2000.
- [17] Vaupel J. W., Manton K. G., Stallard E.: The impact of heterogeneity in individual frailty on the dynamics of mortality. Demography 16 (1979), 439454.
- [18] Wei L. J., Lin D. Y., Weissfeld L.: Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. Journal of the American Statistical Association 84 (1989), 10651073.
- [19] Wienke A.: Frailty Models in Survival Analysis. Habilitation. Martin-Luther-Universität Halle-Wittenberg, 2007. available at <http://sundoc.bibliothek.uni-halle.de/habil-online/>

Contact**Mgr. Jana Füstová**

Faculty of Medicine and Dentistry

Palacky University

Tr. Svobody 8

771 26 Olomouc

Czech Republic

e-mail:jana.furstova@email.cz

Electronic Health Record for Temporomandibular Joint Disorders - Support in Therapeutic Process

R. Hippmann^{1,2}, M. Nagy^{2,3}, T. Dostálová^{1,2}, J. Zvárová^{2,3}, M. Seydlová^{1,2} and E. Feltlová¹

¹Department of Paediatric Stomatology, Second Faculty of Medicine of Charles University, Prague, Czech Republic

²Centre of Biomedical Informatics Prague, Czech Republic

³Institute of Computer Science, Academy of Sciences of the Czech Republic, Prague, Czech Republic, Mathematical Institute AS CR, Prague, Czech Republic

Supervisor: Prof. MUDr. Tatjana Dostálová, DrSc., MBA

Summary

Our experimental work is engaged in creating a new type of Electronic Health Record (EHR) in the field of dentistry, especially for temporomandibular joint (TMJ) disorders (TMD) and the main goal is creation of the comprehensive system, which would be a good basis for the decision support in the TMD therapy.

Our EHR system is based on the Dental knowledge base, which is constructed in the application MUDR (Multimedia Distributed EHR) and then in the MUDRLite application. The MUDR Lite architecture is built on two tiers (relational database (e.g. MS SQL server) and User Interface layer (MUDRLite UI), which is described by an XML configuration file). These applications are creating the real graphical user application DentCross, which is connected with the ASR (Automatic speech recognition) system and TTS (Text-to-speech) module. Special TMD EHR is a part of the Dentcross and is created in the same applications. This part uses the RCD/TMD classification (Research diagnostic criteria for TMD).

We present the Integrated Dental-TMJ EHR system, which allows recording patient's complex medical data in the field of dentistry. It affects the whole orthognathic system with its all components (especially TMD problematic) in the graphical form. The recording and data manipulation are enhanced with the ASR system and TTS module. This application has been constructed with an aim on ease of control and capability of covering medical information and should facilitate decision process in the therapy of TMD.

The integrated Dental - TMJ EHR system represents whole patient's dental information, which is essential for the

treatment plan creation. Connection with the ASR system and TTS module simplifies collecting of this information. Information is also well-arranged in the graphical form. Next step is continuation of clinical testing of the whole system and improvement of the ASR system and TTS module. The disadvantage of this system could be necessity of experience in operating with this system and ASR system functionality in the noisy background.

Keywords: electronic health record, automatic speech recognition, dental cross, temporomandibular joint, temporomandibular joint disorders, structured data entry, dentistry, data model, text-to-speech system, Research Diagnostic Criteria for TMD

1. Introduction

In all fields of medicine the amount of different types of information is increasing. It is highly necessary to have complex information about patient's health status (history, investigation and therapy data). It is also important that this information is always available and stored in a comprehensive form. All these requirements can be fulfilled only by modern computer systems especially by the Electronic Health Record (EHR). Therefore the need of adopting a suitable EHR system (EHR-S) is gradually increasing. However, various studies reveal conflicting data regarding time effectiveness on workflow due to computerization. One of the major issues for physicians is to evaluate potential impact of EHR-S on workflow [1]. Another major issue for the purpose of establishing EHR-S is business-oriented health data distribution in different systems, which

should be integrated to focus on individuals [2]. Generally, EHR-Ss should support continuous, efficient and high-quality integrated healthcare by using structured data entries [3], [4], interoperability [5] and standards [6]. At the beginning of the development of different EHR-Ss the International standard ISO/DTR 20514:2005 was published and defined the EHR as a repository of information regarding the health of a subject of care in a computer process able form, stored and transmitted securely, and accessible by multiple authorized users. The same process of the computerization has been started also in the dentistry. It was performed by various systems. They offered many graphical approaches, but none of them is supported by an automatic speech recognition system. Voice commands usage has been experimented since 1990's [7] as a convenient replacement of a computer control. The necessity of using a human voice to control a computer or another device arises in typical hands-busy environments such as surgery or dentistry [8].

Another challenge is the real complexity of the system. This leads to the creation of the supplementary EHR-S, which is concerned with temporomandibular joint issues and consequently covers the whole orthognathic system.

TMD is an umbrella title for symptoms as temporomandibular dysfunction, cranio-mandibular disorder and mandibular dysfunction. TMD includes a variety of conditions associated with pain and dysfunction of the temporomandibular joint and masticatory muscles [9]. An estimated 20% of the population is affected, with 10% to 20% of those seeking treatment [10], [11].

The main symptoms are pain in the masticatory muscles, limitations or deviation of the mouth opening and TMJ sounds. The etiology of the TMD is multifactorial, therefore detail and complex information is for the therapy indication very important. The diagnosis of TMD is commonly based on the presenting signs and symptoms. For the classification among clinicians and researchers The Research Diagnostic Criteria for the Temporomandibular Disorders (RDC/TMD) are used. According to the mentioned facts, our main goal is to develop an EHR focused on the TMD, which will lead to the decision support system in the therapy of TMD.

2. Methods

2.1 MUDR EHR, Dental Knowledge database and MUDRLite

The EuroMISE Centre has a long tradition in research and development of EHR-Ss, which started in the year 2000. The research resulted in a pilot EHR application called MUDR (Multimedia Distributed EHR). The main requirement for the system was the storage of structured data combined with free text with the possibility of dynamic extension and modification of the set of collected attributes without any change of the database structure. The MUDR EHR was based on 3-tier architecture, consisting of a database layer, application layer and user interface. The set of definitions of collected attributes and relations between them is stored in a directed graph structure called "knowledge base".

The modelling tools were developed as a part of MUDR EHR and were employed in this project to create a Dental Knowledge Base. This knowledge base was created in cooperation between IT specialists and dentists in years 2004 and 2005. It comprises of approximately 1000 concepts that cover all possible situations that we can find in the patient's dental status and basic facts about the patient as well.

Since the MUDR EHR was designed for large scale enterprises, which was in some cases a limiting factor, another EHR-S was developed at the EuroMISE Centre in 2005. This EHR-S was related to MUDR EHR, hence named MUDRLite [12] and

was developed with the emphasis on the small practice needs. The MUDRLite architecture is based on two tiers. The first one is a relational database (e.g. MS SQL server) and the second one is a MUDRLite User Interface layer (MUDRLite UI). The database schema corresponds to the particular needs of the target domain and therefore varies significantly in different environments. The core of MUDRLite MUDRLite Interpreter is able to handle various database schemas. This feature often simplifies the way of importing old data stored in other databases or files. The visual aspects as well as the behaviour of the MUDRLite UI are completely described by an XML configuration file. The end-user can see a set of forms with various controls placed on them by appropriate XML elements.

MUDRLite operates as a command interpreter; it processes the instructions encoded in the so called MLL (MUDRLite Language) [13] based on XML and manipulates the database layer as well as the visual aspects and behaviour of the MUDRLite user interface. The pilot application was prepared for dentistry [8] and as a next step for the TMD, which will be mentioned below.

2.2 DentCross component, ASR

EHR systems MUDR and MUDRLite were usable for the general clinicians. In the field of dentistry it is common to have information presented in the graphical form. For this reason and for better user acceptance an advanced component representing the dental cross was developed, which is clearly organized in the graphical form. The DentCross component is implemented as a stand-alone library DentCross.dll, completely developed for the .NET Framework platform. This user-defined component is inserted by the custom element of the MLL (named "custom") with the following mandatory attributes: "dll" specifying the name of the assembly where the component is implemented and "class" specifying the name of the main class of the included component.

The user interface for collection of the TMD data is a user form defined completely in MLL. It is integrated in the same application as the DentCross is, thus

creating a comprehensive EHR in dentistry [14].

Real interactivity was achieved by adoption of the system for the automatic speech recognition and the text to speech module. The automatic speech recognition (ASR) engine is described in more detail in [15] and [16].

2.3 RDC/TMD

In the specialized part of the EHR aimed on the TMD we have used basic schema for the anamnesis. For the TMD classification we have used parts of RCD/TMD, which applies the dual-axis system to diagnose and classify patients with TMD [17]. The first axis is divided into 3 groups of commonly occurring TMDs:

1. muscle disorders, including myofascial pain,
2. disk displacement with or without reduction,
3. arthralgia, arthritis and arthrosis.

The second axis in this nomenclature is using the 31-items questionnaire, used to evaluate behavioural, psychological and psychosocial factors. We have modified these investigated factors and we have created our own questionnaire for evaluation of the quality of life.

3. Results

A knowledge modelling tool from the MUDR EHR application suite was used to create a specialized TMD Knowledge Base. This knowledge base contains more than 80 concepts (Fig. 1) and extends the former Dental Knowledge Base.

In order to create MUDRLite forms, concepts from the TMD Knowledge Base had to be transformed in a relational database model (Fig. 2). The user interface and application for TMD was encoded into MLL in MUDRLite EHR, thus creating user forms interconnected with a database layer.

The basic part of our interactive EHR in dentistry is composed of the DentCross component with the ASR system, which is built on concepts originating from the Dental Knowledge Base and uses the MUDRLite EHR as a hosting system. The specialized TMD EHR contains concepts defined in the TMD Knowledge Base.

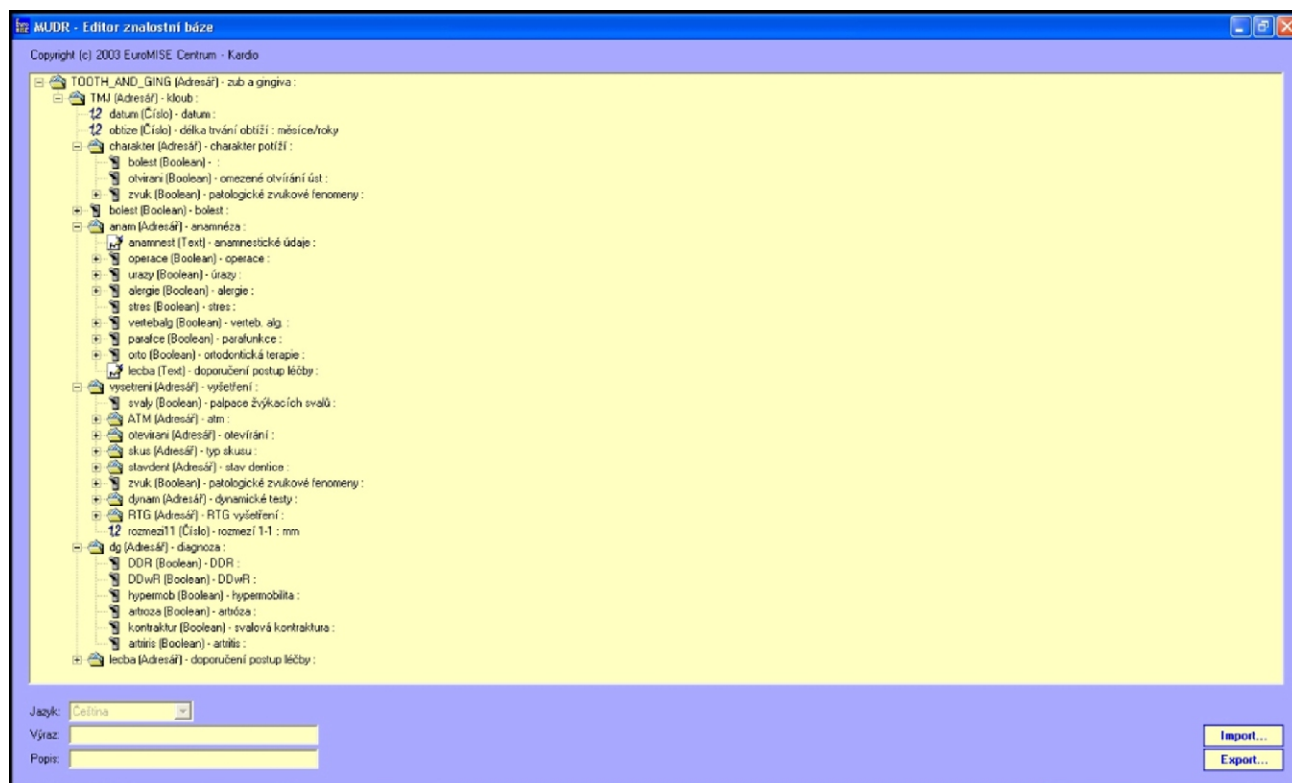


Fig. 1. TMD knowledge base.

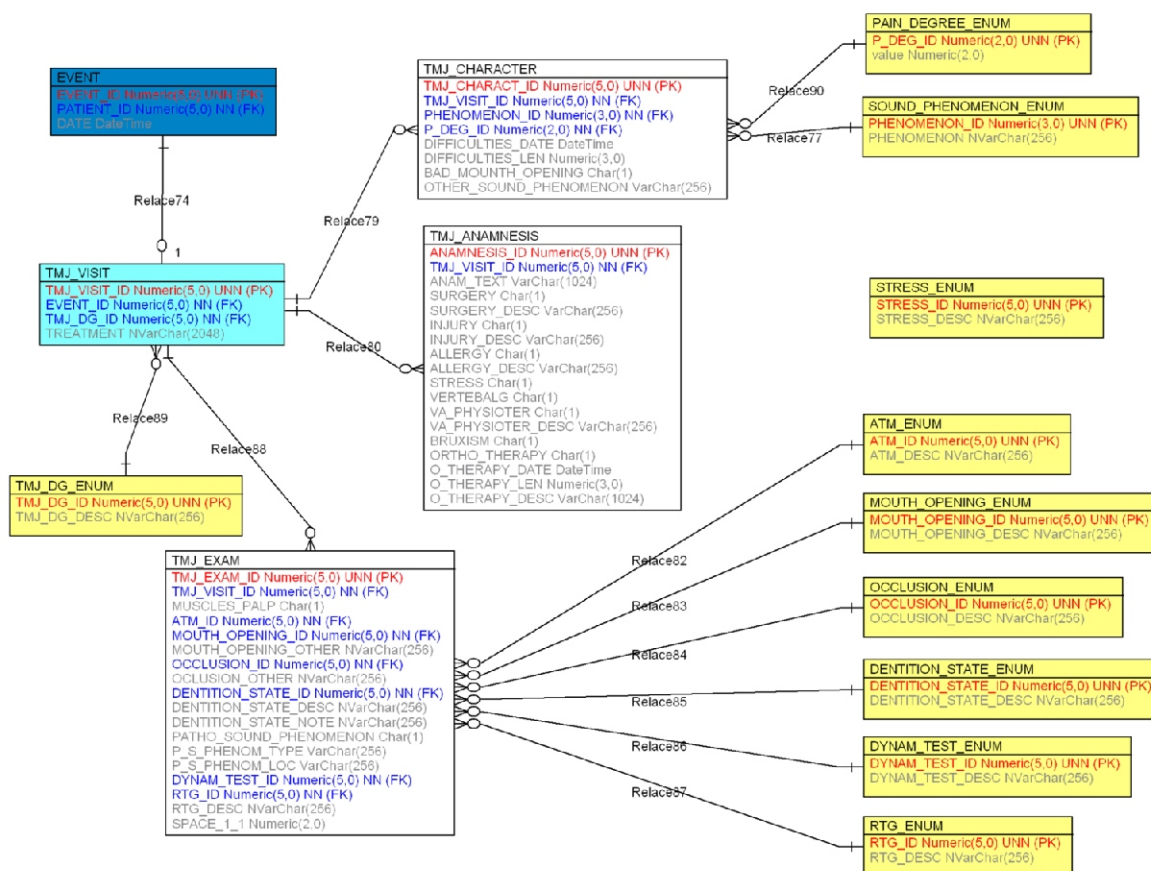


Fig. 2. Data model for TMD EHR and its tree-structure.

It is initiated from the main window of the Integrated Dental-TMJ EHR (DentCross Component, ASR, TTS and TMD part) and has its own user interface.

This application consists of 3 main parts (Fig. 3). In the first part we can find basic patient facts (name, id number, date of birth etc.), which helps in orientation. The second part is the list of events. They are sorted by the dates and by opening these dates we can reach and open records connected with one particular patient visit. The third part consists of the subjective patient problems, anamnestic data, results of investigation, dg statement and therapy recommendations. This part is the most important and represents all data about the particular patient.

It is very important to know, what circumstances lead a patient to visit the specialist and what patient's personal feelings are. The clinician can recognize many important facts from the patient description. We can record a type and character of difficulties, their duration and closer description. In the anamnestic section we can record basic personal anamnestic data and different conditions important to the TMD (parafunctions, stress, orthodontic treatment).

Probably the most important is the section with investigation modalities. Investigation is in the TMD classified in aspection (deformities, status of dentition, mouth opening (Fig. 4) etc.), palpation (masticatory muscles, TMJ) and auscultation (TMJ sounds). The last part in the TMD investigation playing the crucial role is X-ray imaging (ortopantomographic image, computer tomography (Fig. 5)) and magnetic resonance. Subscription and presence of these images can be also recorded. In the diagnosis classification we have used above mentioned RDC/TMD. We can record also therapy recommendation, which can be changed at the end of each visit. This application serves as a complex information source for the TMD treatment and is a part of the whole system for the support in decision in the therapy of TMD, which leads in the final treatment plan:

1. implementation to the interactive DentCross component - dental examination and anamnestic data,
2. training of opening of the mouth in the axis, without deviation,
3. stabilization splint,
4. computer-aided face bow analysis (Fig. 6),
5. prosthetic treatment,

6. finalization of the treatment and soft splint,

7. recall and treatment evaluation by the patient.

4. Discussion

The presented DentCross component especially with the specialized part for the TMD and ASR system seems to be highly suitable for the field of dentistry. Its graphical form offers a good overview during patient's investigation and treatment. Voice commands and the system for ASR enable control of the data recording without other person's assistance. Another advantage is the possibility of data recording without a traditional control by a mouse and keyboard. This is very important especially in the dentistry and medicine in general. A hand with the glove is in the permanent contact with the potentially infectious material. The DentCross component is used in the dental office in the Department of paediatric stomatology, Faculty hospital Motol and was used in the Department of biological analysis Institute of Criminalistics in Prague for dead corpses' identification [8].

The screenshot displays the 'Stomatologický pacient - TMJ' application window. It features a top bar with patient identification fields (Name, Surname, ID, Birth Date, Version) and a 'Zavřít' button. Below this is a 'Seznam TMJ událostí' (TMJ Event List) table with columns for ID and Datum. The table lists five events from 2007 to 2008. To the right, the 'Situace TMJ' (TMJ Situation) section contains various input fields for clinical data, including 'Datum', 'Trvání potíží', 'Omezené otevírání úst', 'Bolest', 'Zvukové fenomény', 'Anamnéza', 'Operace', 'Úrazy', 'Alergie', 'Stres', 'Vertebrální', 'Rehabilitace', 'Bruzismus', 'Orto. terapie', 'Dg', and 'Dop. chir. vrb.'. The 'Výšetření' (Examination) section includes fields for 'Palpce žvýkacích svalů', 'ATM', 'Otevírání', 'Typ skusu', 'Stav dentice', 'Pops', 'Poznámka', 'Patologické zvukové fenomény', 'Typ', 'Lokalizace', 'Dynamické testy', 'RTG vyšetření', 'Pops', and 'Rozmezí 1-1 (mm)'. At the bottom, there are buttons for 'Nový' and 'Uložit', and a field for 'Doporučený postup léčby'.

Fig. 3. TMD EHR application.

The specialized TMD part of the Integrated Dental - TMJ EHR extends the contemporary EHR system in the dentistry. It consists of basic anamnestic data, investigation data, diagnose statement and therapy recommendation. This application was used in the special office at the Faculty hospital Motol for the therapy of the TMD and was a part of the system for the decision support in the therapy of TMD. It allows a complex overview during investigation, treatment planning and it also saves some time.

The future work is to improve the Integrated Dental-TMJ EHR and create an automatic system built on the anamnestic data signs and symptoms to make a diagnosis and prepare a therapeutic plan. This would facilitate the whole therapy process of the multifactorial conditioned TMD.

Acknowledgments

This work was supported by the Grant 9991-3 IGA Ministry of Health CR, project FI-IM5/142, 1M06014 Ministry of Education, Youth and Sports and SVV-2010-265513.

References

- [1] Shabbir S.A., Ahmed L.A., Sudhir R.R., Scholl J., Li Y.C., Liou D.M.: Comparison of documentation time between an electronic and a paper-based record system by optometrist at an eye hospital in south India: A time.motion study. *Comput Methods Programs Biomed*, 2010 May 10.
- [2] Liu D., Wang X., Pan F., Yang P., Xu Y., Tang X., Hu J., Rao K.: Harmonization of health data at national level: a pilot study in China. *Int J Med Inform*, 2010 Jun; 79(6): 450-8.
- [3] Los R.K., van Ginneken A.M., van der Lei J.: OpenSDE: a strategy for expressive and flexible structured data entry. *Int J Med Inform* 2005; 74:481-490.
- [4] van Ginneken A.M.: The computerized patient record: balancing effort and benefit. *Int J Med Inf* 2002; 65:97-119.
- [5] Bakker A.: Access to EHR and access control at a moment in the past: A discussion of the need and an exploration of the consequences. *Int J Med Inform* 2004; 73:267-270.
- [6] Blobel B.: Advanced EHR Architecture Promises or Reality. *Methods Inf Med* 2006; 1: 95-101.
- [7] Grasso M.A.: Automated speech recognition in medical applications. *MD Computing* 1995; 12: 16-23.



Fig. 4. TMJ - data collection. Measurement of the mouth opening.

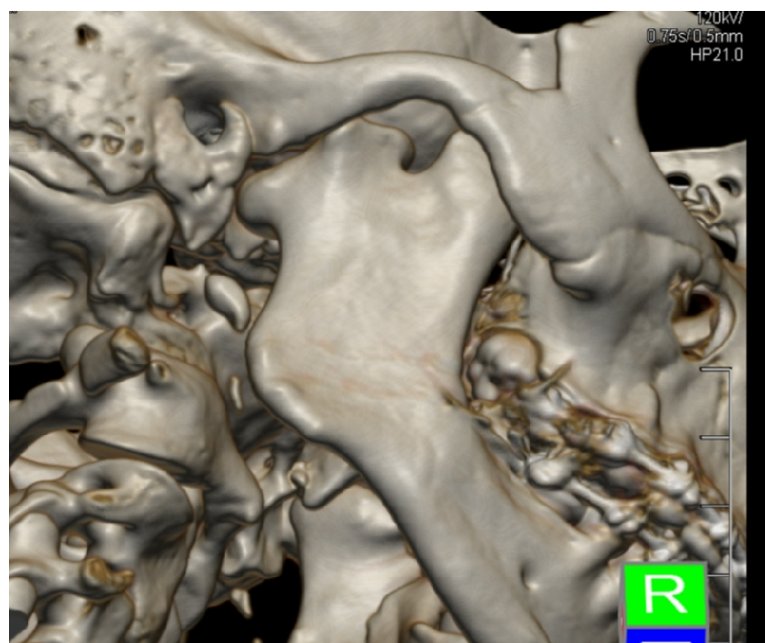


Fig. 5. TMJ - 3D-CT computer image analysis. In this picture is presented TMJ ankylosis.

Bite Fork: Position 2			
Right Joint		Left Joint	
HCN/HCI	05.7 (CE)	HCN/HCI	-0.9 (CE)
Bennett	04.0	Bennett	04.0
Iss	00.0	Iss	00.0
Shift Angle	-17.9	Shift Angle	-20.0
Anterior Guidance			
Right	18.5		
Middle	14.7		
Left	16.0		
		Exit	Reset

Fig. 6. TMJ - computer face bow analysis.

- [8] Zvarova J., Dostalova T., Hanzlicek P., Teuberova Z., Nagy M., Pies M., Seydlova M., Eliasova H., Simkova H.: Electronic health record for forensic dentistry. *Methods Inf Med* 2008; 47: 8-13.
- [9] Di Fabio R.P.: Physical therapy for patient with TMD: a descriptive study of treatment, disability and health status. *J Orofac Pain.* 1998;12:124-135.
- [10] Pedroni C.R., De Oliveira A.S., Guaratini M.I.: Prevalence study of signs and symptoms of temporomandibular disorders in university students. *J Oral Rehabil.* 2003; 30: 944-950.
- [11] Taylor M., Suvinen T., Reade P.: The effect of Grade IV distraction mobilization on patients with temporomandibular pain-dysfunction disorder. *Physiotherapy Theory and Practice.* 1994;10:129-136.
- [12] Hippmann R., Dostalova T., Zvarova J., Nagy M., Seydlova M., Hanzlicek P., Kriz P., Smidl L., Trmal.: Voice supported electronic health record for temporomandibular joint disorders. *Methods of information in medicine*, 2010; 49:168-172
- [13] Spidlen J., Hanzlicek P., Zvarova J.: MUDRLite – health record tailored to your particular needs. *Studies in health technology and informatics* 2004; 105: 202-9.
- [14] Hippmann R., Dostalova T., Zvarova J., Nagy M., Seydlova M., Hanzlicek P., Kriz P., Smidl L., Trmal.: Voice supported electronic health record for temporomandibular joint disorders. *Methods of information in medicine*, 2010; 49:168-172.
- [15] Nagy M., Hanzlicek P., Zvarova J., Dostalova T., Seydlova M., Hippmann R., Smidl L., Trmal J., Psutka J.: Voice-controlled Data Entry in Dental Electronic Health Record. *Studies in Health Technology and Informatics* 2008; 136: 529-534.
- [16] Psutka J., Muller L., Matousek J., Radova V.: *We speak Czech with a computer* Praha, Academia, 2006.
- [17] Ali H.N.: Diagnostic criteria for temporomandibular joint disorders: a physiotherapist's perspective. *Physiotherapy.* 2002; 88:421-426.

Contact

MUDr. Radek Hippmann

Department of Paediatric Stomatology,
Second Faculty of Medicine of Charles
University
V úvalu 84
150 06 Prague 5
Czech Republic
e-mail: r.hip@seznam.cz

Comparison of Texture Classifier and Human Observer in Diagnosis of Autoimmune Thyroiditis, Observer Variability Evaluation

Š. Holinka¹, D. Smutek¹

¹3rd Department of Medicine, 1st Medical Faculty, Charles University in Prague, Czech Republic

Supervisor: Doc. MUDr. Ing. Daniel Smutek, Ph.D.

Summary

The objective has been to compare success of the texture classifier and a human observer in diagnosis of the autoimmune thyroiditis from B-mode ultrasound images and to determine inter- and intra-observer variability.

The data set of 161 subjects was classified by four human observers and by the Bayes classifier based on the texture features to three classes (healthy, border state, autoimmune thyroiditis).

Two observers had a higher success rate when classifying the healthy class (74.4% and 83.3%), the other two observers classified better cases with autoimmune thyroiditis (59.0% and 77.4%). The classifier gave the relatively high and balanced success rate for both classes (100.0% for healthy and 87.5% for thyroiditis). The different observers' success rates resulted in the high inter-observer variability, showing only a fair agreement among the human observers. There was no significant difference among human observers in the intra-observer variability.

Due to the fair agreement among observers in the diagnosis of autoimmune thyroiditis from ultrasound images and good results of the classifier, the best way in establishing diagnosis is computer-aided diagnosis combined with observers' clinical experience.

Keywords: thyroid gland, autoimmune thyroiditis, ultrasound image, B-mode ultrasound, texture analysis, computer-aided classification, inter-observer variability, intra-observer variability, Kappa Statistics, weighted Kappa.

1. Introduction

Autoimmune thyroiditis, one of the most frequent diseases of the thyroid gland, is a chronic inflammation of the thyroid parenchyma [1]. The inflammation of the

gland causes diffuse changes in the structure of the thyroid tissue. These changes can be detected by ultrasound imaging which is the most widely used diagnostic and monitoring tool for this disease. Thus, the method is suitable as a complementary examination method next to the diagnosis done from clinical examination using immunologic, hormonal and metabolic analyses of blood samples, and from cytological examination using fine-needle aspiration biopsy. However, the assessment of the diffuse processes is difficult [2], [3] and, in daily practice, the diagnosis from B-mode ultrasound images is made qualitatively from the size of the gland being examined, its perfusion, and the structure and echogenicity of its parenchyma. A physician uses clinical experience without giving any quantifiable indices.

This subjective evaluation of the ultrasound image texture is a reason why computer-aided methods for an automatic diagnosis were proposed. In the context, the suitable texture features were founded [4], [5], and a successful classifier for a semi-automatic diagnostic method of autoimmune thyroiditis from B-mode ultrasound images constructed [6].

Consequently, the endeavours in comparing computer-aided detection of various disorders with human reading of medical images have appeared. For example the study from the field of screening mammography [7] or the field of pletysmography [8].

In this paper, we aimed to compare results of human observers (endocrinologists) in a diagnostics of autoimmune thyroiditis from B-mode ultrasound images with results achieved by the classifier described in [6]. Moreover, acquired data were evaluated on inter- and intra-observer variability of human observers.

2. Materials and methods

The data set consists of B-mode ultrasound images of 161 patients (subjects) from the Department of Endocrinology who were referred ultrasound examination of the thyroid gland. The subjects were consecutively screened to the study during three months enrolment period. The subjects with local changes such as nodules were excluded. For each subject from 4 to 8 B-mode ultrasound images were scanned in both longitudinal and transverse planes (see Figure 1) by the ultrasound system EnVisor M2540A equipped with an 8MHz linear probe and were saved in the DICOM format.

The data set were divided to the three classes according to the diagnosis confirmed by a clinical examination: H (normal gland, 26 subjects), BS (borderline state between healthy and inflamed tissue, 14 subjects) and AT (autoimmune thyroiditis, 121 subjects). Moreover a fourth class "Evaluation impossible" was added, which could be used in cases where a human observer was not able to conclude the diagnosis.

Four observers (A, B, C, D) are all endocrinologists with at least 10 years of experience in the ultrasound examination of the thyroid gland. They evaluated the anonymized ultrasound data by using a web-application, developed for this purpose, without any knowledge about the previous diagnosis and clinical results. The observers were instructed to choose one of the classes (H, BS, AT and "Evaluation impossible") for each subject displayed in the random order in the web application. Time for evaluation was not limited. In general, the observers executed this during three weeks with five second per subject on average. Each subject was evaluated three times (in three rounds) by each of four observers completely.

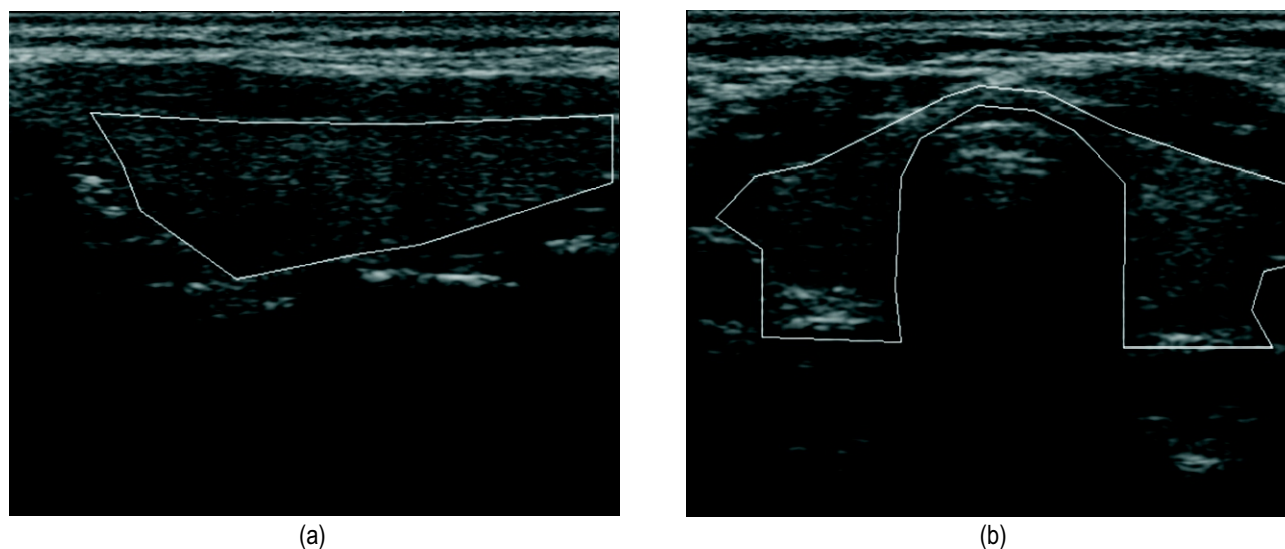


Fig. 1. Ultrasound images of thyroid gland scanned for subjects with delineated region of interest: (a) longitudinal scan of inflamed tissue; (b) transverse scan of healthy tissue.

Simultaneously the ultrasound data were evaluated by the classifier (CL) designed and verified in the last project [6]. The classifier is based on the Bayesian decision theory and used the texture features designed by Muzzolini and Haralick [9], [10]. As different ultrasound scanners give different numerical values for same texture features [5], new feature selection was done for the ultrasound scanner used in this study. As the most suitable feature was selected Muzzolini's spatial texture feature, which represented a gradient magnitude of the image texture [9]. The data set was split into a training set and an independent test set (see Table 1) to allow classifier training to the ultrasound image texture acquired from the used scanner.

At first, the success of observers and the classifier in diagnosing to the correct classes were compared. Secondly, inter- and intra-observer variability of human observers were measured by kappa statistic [11]. The kappa values are usually obtained from the interval $<0,1>$ and are interpreted as a poor agreement for $\kappa = 0$ and a perfect agreement for $\kappa = 1$. The inter-observer variability was quantified separately for the each evaluation round

by multiple-reader kappa κ_m and weighted kappa κ_w with quadratic weights and computed with 95% confidence limit. Weighted kappa is especially developed to allow evaluate importance of disagreements when the observer decision is made over multiple categories (three classes H, BS, AT) [12].

The intra-observer variability was quantified also by using the multiple-reader kappa κ_m , where the calculation was done on three evaluation rounds per observer instead of the calculation done on group of observers.

3. Results

3.1 Comparative study

The success in the evaluation of the whole data set by observers (A, B, C, D) and by the classifier (CL) is shown in Table 2.

Each classification of each subject was supposed to be an independent random experiment. Hence, total evaluations number given by the each observer was 483 in accordance with 161 subjects independently classified three times.

The success rate of the classifier was determined on the independent test set including only 52 subjects because remaining subjects were used to train the classifier. And there was no need to do the evaluation three times because of the 100% reproducibility of the classifier results.

In the Table 2, the column "Not classified" is a number of subjects where observers used classification to the class "Evaluation impossible". The classification success is shown as a number of correct evaluations as well as frequencies, which are represented as a ratio of the number of correct evaluations and the number of evaluations in separate classes. For example for the observer A:

$$\begin{aligned} \text{H:} & \quad 33 / (3 \cdot 26) = 0,423 \\ \text{BS:} & \quad 28 / (3 \cdot 14) = 0,667 \\ \text{AT:} & \quad 214 / (3 \cdot 121) = 0,590 \end{aligned}$$

The major misclassification expresses subjects from the class H classified incorrectly to the class AT and vice versa. The minor misclassification expresses the other incorrect classifications (from BS to H or AT and vice versa).

Tab. 1. Available data.

	class H	class BS	class AT
training set	18	10	81
independent test set	8	4	40

Tab. 2. Success in diagnosis of observers (A,B,C,D) and classifier (CL) in three separate classes H, BS, AT. Success is shown as a number of correct evaluations and frequencies in parentheses.

Observer	Success in separate classes			Overall success	Major misclassif.	Minor misclassif.	Not classified
	H	BS	AT				
A	33 (0.423)	28 (0.667)	214 (0.590)	275 (0.569)	16 (0.033)	190 (0.393)	2
B	58 (0.744)	22 (0.524)	173 (0.477)	253 (0.524)	47 (0.097)	165 (0.342)	18
C	13 (0.167)	25 (0.595)	281 (0.774)	319 (0.660)	13 (0.027)	151 (0.313)	0
D	65 (0.833)	16 (0.381)	139 (0.383)	220 (0.455)	75 (0.155)	169 (0.350)	19
CL	8 (1.000)	4 (1.000)	35 (0.875)	47 (0.904)	2 (0.038)	3 (0.058)	0

Tab. 3. Contingency tables displaying results of the classification done by human observers and the classifier (leave-one-out classification process on the training data set and the classification on the independent test set). Columns of contingency tables represent true classes, each row represents classification results and the main diagonal represents correct classifications.

Observer A					Observer B				
	H	BS	AT	Not classif.		H	BS	AT	Not classif.
H	33	38	7	0	H	58	18	1	1
BS	9	28	5	0	BS	16	22	4	0
AT	9	138	214	2	AT	46	127	173	17
Observer C					Observer D				
	H	BS	AT	Not classif.		H	BS	AT	Not classif.
H	13	53	12	0	H	65	13	0	0
BS	1	25	16	0	BS	24	16	1	1
AT	1	81	281	0	AT	75	131	139	18
Classifier - LOO on training set					Classifier – independent test set				
	H	BS	AT			H	BS	AT	
H	18	0	0		H	8	0	0	
BS	3	6	1		BS	0	4	0	
AT	5	8	68		AT	2	3	35	

The minor and the major misclassification frequency are ratios of the number of incorrect classifications and total evaluations per observer (3 * 161) or the classifier (52).

Classification results of human observers and classifier are shown in contingency tables (Table 3).

3.2 Inter- and intra-observer variability

The inter- and intra-observer variability of human observers were examined, see Table 4, 5 and 6. The column N appeared in the following tables denotes the number of all classified subjects except those classified to the class "Evaluation impossible".

Three separate evaluation rounds of 161 subjects per observer were discriminated. The inter-observer variability was quantified separately for each round. In Table 4 multiple-reader kappa κ_m values represented overall agreement among observers. The amount of identically classified subjects by all observers is also mentioned. Weighted kappa κ_w values were evaluated to obtain detailed relation dependencies for pairs of observers.

Tab. 4. Inter-observer (A,B,C,D) agreement, multiple-reader kappa κ_m separately for three evaluation rounds.

Observers	N	All four observers agreed	κ_m
A,B,C,D (r.1)	152	63 (0.416)	0.448
A,B,C,D (r.2)	154	39 (0.253)	0.264
A,B,C,D (r.3)	147	31 (0.211)	0.258

Tab. 5 Inter-observer (A,B,C,D) agreement between paired observers. Class agreement for H, BS and AT and weighted kappa κ_w separately for three evaluation rounds.

;Pair of observers	N	Class agreement			κ_m	Confidence limits
		H	BS	AT		
A&B (r.1)	156	0.593	0.536	0.795	0.670	0.622 - 0.777
A&B (r.2)	158	0.333	0.465	0.707	0.544	0.453 - 0.635
A&B (r.3)	151	0.585	0.565	0.726	0.647	0.667 - 0.737
A&C (r.1)	160	0.533	0.682	0.823	0.698	0.614 - 0.781
A&C (r.2)	161	0.286	0.593	0.790	0.535	0.422 - 0.648
A&C (r.3)	160	0.174	0.474	0.754	0.499	0.400 - 0.598
A&D (r.1)	155	0.480	0.442	0.824	0.673	0.594 - 0.751
A&D (r.2)	155	0.281	0.397	0.720	0.544	0.462 - 0.627
A&D (r.3)	153	0.520	0.450	0.600	0.569	0.485 - 0.653
B&C (r.1)	156	0.429	0.536	0.823	0.675	0.570 - 0.751
B&C (r.2)	158	0.167	0.212	0.671	0.463	0.383 - 0.543
B&C (r.3)	151	0.087	0.226	0.627	0.402	0.320 - 0.484
B&D (r.1)	152	0.706	0.455	0.779	0.717	0.634 - 0.800
B&D (r.2)	154	0.714	0.527	0.720	0.698	0.614 - 0.782
B&D (r.3)	147	0.755	0.605	0.675	0.710	0.629 - 0.792
C&D (r.1)	156	0.286	0.385	0.759	0.521	0.423 - 0.619
C&D (r.2)	155	0.139	0.180	0.618	0.388	0.306 - 0.471
C&D (r.3)	153	0.069	0.175	0.463	0.298	0.225 - 0.370

Tab. 6. Intra-observer (A,B,C,D) agreement, multiple-reader kappa κ_m for individual observers.

Observers	N	All three rounds agreed	κ_m
A	160	94 (0.588)	0.534
B	151	89 (0.589)	0.578
C	161	108 (0.671)	0.545
D	151	80 (0.530)	0.525

Values are shown in Table 5 with class agreement between two observers which is calculated as a number of correct evaluations done by both observers divided by a number of subjects classified to this class by both observers [10], [11].

The intra-observer variance (Table 6) represents agreement in evaluation of each subject in three rounds for each observer. The different values in column N are given by omitting subjects marked as "Evaluation impossible".

4. Discussion and Conclusion

In the comparative study the evaluation success of B-mode ultrasound images by human observers and the classifier were compared. Observers A and C and the pair of observers B and D achieved similar success rates. Thus observers were divided into two groups. Observers A and C achieved the high success rate in the

classification to the class AT (59.0% and 77.4% respectively). However, the observer C also classified several healthy subjects (class H) to classes representing the positive disease, i.e. false-positive rate 12 from 13 major misclassifications were done from the class H to the class AT. The observer A major misclassification was balanced 7 from 16 were done from the class H to the class AT (see Table 3).

To the contrary, observers B and D classified more successfully to the healthy class (74.4% and 83.3%) and used more the possibility not to classify unclear subjects (18 and 19 subjects respectively). Further, observers B and D classified with false negative rate. They misclassified 47 and 75 subjects from AT class as healthy (class H) respectively, as shown in Table 3.

The classifier was verified by leave-one-out classification on the training set, where good results (84.4%) were acquired. Classification success on 52 subjects from the independent test set was relatively high and balanced for all three considered classes (100.0% for H, 100.0% for BS and 87.5% for AT). There were 2 major misclassifications both from AT to H (see Table 3), which represent the higher false negative rate. As mentioned in Introduction, the observer's (physician's) decision is based mainly on subjective evaluation of the structure and echogenicity of parenchyma in ultrasound images and their clinical experience, whereas the classifier evaluation is based on quantifiable indices. Thus, the computer-aided diagnosis of autoimmune thyroiditis combined with physician clinical experience would be the best way.

In inter- and intra-observer variability study, it was noted that in the first evaluation round observers agreed more ($\kappa_m = 0.448$) than in two following rounds ($\kappa_m = 0.264$ and 0.258), see Table 4. The reason could have been that the observers in the first round evaluated data set with considerable interest and higher attention than in two following rounds.

There can be seen in Table 5, that the agreement in pairs of the observer C against B and D, represented by values of κ_m , is lower than in pairs of observers A, B and D against each other. The reason is in the substantial false-positive rating in the data set classification by observer C (see Table 2) against false-negative rating done by observers B and D. Relationships of observers A, B and D are relatively moderate up to substantial. The best agreement is given in pairs A against B and B against D, which is probably caused by the observer B, who made decisions on the data set with the success against to clinical examinations somewhere between the classification success of observers A and D (Table 2).

The intra-observer variability quantified by multi-reader kappa (κ_m) values is compared in Table 6. There is no significant difference among observers (κ_m : 0.534, 0.578, 0.545 and 0.525). When we consider kappa values from the interval $<0,1>$ where $\kappa_m = 0$ means the poor agreement and $\kappa_m = 1$ means the perfect agreement, all four observers in our study achieved the moderate agreement which

also means the moderate intra-observer variability.

Acknowledgement

This work has been supported by the Grant Agency of Charles University in Prague, Czech Republic, grant number 119607 and by the grant SVV-2010-265 513.

The authors wish to thank endocrinologists Prof. Dr. Michal Kršek CSc., Dr. Jan Jiskra PhD. and Dr. Petr Sucharda in the Department of Endocrinology of Charles University in Prague for their effort at evaluation of ultrasound images.

References

- [1] Warftsky L., Ingbar S.H. : Disease of the thyroid. In Harrison's Principles of Internal Medicine, page 1712. McGraw-Hill, New York, 12th edition, 1991.
- [2] Simeone F.J., Daniel G.H., Müller P.R. et al. : High-resolution real-time sonography. Radiology, 155:431439, 1985.
- [3] Solbiati L., Volterrani L., Rizzatto G., et al. : The thyroid gland with low-uptake lesions: Evaluation by ultrasound. Radiology, 155:187196, 1985.
- [4] Šára R., Smutek D., Sucharda P., Svačina Š. : Systematic construction of texture features for Hashimoto's lymphocytic thyroiditis recognition from sonographic images. In S. Quaglini, P. Barahona, and S. Andreassen, editors, Artificial Intelligence in Medicine, LNCS, Berlin-Heidelberg, Germany. Springer, 2001.
- [5] Smutek D., Šára R., Sucharda P., Tjahjadi T., Švec M. : Image texture analysis of sonograms in chronic inflammations of thyroid gland. Ultrasound in Medicine and Biology, 29(11):1531 1543, 2003.
- [6] Holinka Š., Šára R., Smutek D. : Relation Between Structural Changes in B-mode

Ultrasound Images of Thyroid Parenchyma and the Presence of Thyroid Antibodies in Blood Sample. Machine Graphics Vision, 18(1):67-82, 2009.

- [7] Georgian-Smith D., Moore R.H., Halpern E. : Blinded Comparison of Computer-Aided Detection with Human Second Reading in Screening Mammography. American Journal of Roentgenology, 189:11351141, 2007.
- [8] Paez S., Seiden D., Kiel M., Chediak A. : A comparison of human and computer apnea/hypopnea detection from respiratory waveforms derived by respiratory inductive plethysmography. Sleep Research, 25: 521, 1996.
- [9] Muzzolini R., Yang Y.-H., Pierson R. : Texture characterization using robust statistic. Pattern Recognition, 27(1):119134, 1994.
- [10] Haralick R.M., Sapiro L.G. : Computer and vision. vol. 1. Reading: Addison-Wesley Publishing Company, 1993: 453 506.
- [11] Fleiss J.L., Levin B., Paik M.C. : Statistical methods for rates and proportions. John Wiley, New York, third edition, 2003.
- [12] Kundel H.L., Polansky M. : Measurement of observer agreement. Radiology, 228(2):303308, August, 2003.

Contact

MUDr. Štěpán Holinka

3rd Department of Medicine,
1st Medical Faculty, Charles University
in Prague
U Nemocnice 1
128 08 Prague 2
Czech Republic
e-mail: stepanholinka@gmail.com

On the Relation between Generalized Entropy and the Bayes Decision Error

M. Horáček¹

¹Centre of Biomedical Informatics, Institute of Computer Science, Academy of Sciences of the Czech Republic

Supervisor: Prof. RNDr. Jana Zvárová, DrSc.

Summary

We deal with the relation between the generalized entropy (f -entropy, a family of functions that include several biodiversity measures) of a discrete random variable and the minimal probability of error (Bayes error) when the value of this random variable is estimated. Namely the tightness of their relation is studied.

Morales and Vajda [1] recently introduced a measure called average inaccuracy that aims to quantify the tightness of the relation between the posterior Bayes error and the power entropies. It is defined as a standardized average difference between the upper and the lower bound for the posterior Bayes error under given entropy. Their concept can be generalized to any strictly concave f -entropy and used to evaluate its relation to the Bayes probability of error. However, due to a complex form of the formula of the average inaccuracy, it is difficult to compare the average inaccuracies of most f -entropies analytically.

We propose a smooth approximation of the lower bound for the posterior Bayes error under given f -entropy that simplifies the formula of the average inaccuracy. We show that under this approximation, the quadratic entropy has the tightest relation to the posterior Bayes error among f -entropies.

The quadratic entropy has the tightest relation to the posterior Bayes error (in the sense described in this paper) than the Shannon's entropy and other functions that belong to the family of f -entropies, like Emlen's index, Ferreri's index and Good's index.

Keywords: generalized entropy, f -entropy, Bayes error, average inaccuracy, power entropy, quadratic entropy, Shannon's entropy, Emlen's index, Ferreri's index, Good's index

1. Introduction

Let X and Y be discrete random variables with values in finite sets $X = (x_1, \dots, x_r)$ and $Y = (y_1, \dots, y_s)$. Let Z be a random variable that is absolutely continuous with respect to the Lebesgue measure with probability distribution P_z on some set $Z \subseteq \mathbf{R}$. We denote $P(x) = P(X = x)$, the joint distribution of X and Y as $p(x, y) = P(X = x, Y = y)$, the conditional distribution of X under given $Z = z$ as $p(x|z) = P(X = x|Z = z)$ etc., i.e. the arguments in the probability distribution p denote the space and random variable which they refer to.

The uncertainty associated with a random variable X is traditionally measured by the Shannon's entropy

$$H(X) = - \sum_{x \in X} p(x) \ln p(x)$$

with convention $0 \ln 0 = 0$. It is intuitively related to the minimal probability of error (the prior Bayes error)

$$e(X) = 1 - \max_{x \in X} p(x)$$

when the value of X is estimated. The conditional Shannon's entropy

$$\begin{aligned} H(X | Z) &= \int_Z H(X | Z = z) dP_z(z) = \\ &= - \int_Z \sum_{x \in X} p(x | z) \ln p(x | z) dP_z(z) \end{aligned}$$

and the average minimal posterior probability of error (the posterior Bayes error)

$$e_B(X | Z) = 1 - \int_Z \max_{x \in X} p(x | z) dP_z(z)$$

are related in a similar manner. Fano [2] and Kovalevsky [3] were among the first who studied the relationship between the entropy and the Bayes error. They derived attainable bounds for Shannon's entropy $H(X)$ under given prior Bayes error $e(X)$. Since these bounds are continuous and strictly increasing functions of the Bayes error e on the interval $[0, (r-1)/r]$, the bounds for the prior Bayes error under given entropy are well defined. Fano's and Kovalevsky's findings were later broaden to more general classes of entropies by various authors, namely to all Schur-concave entropies by Vajda and Vašek [4]. Vajda and Vašek also retrieved corresponding bounds for conditional entropies and the posterior Bayes error.

Recently, Morales and Vajda [1] studied tightness of the relationship between the prior and posterior Bayes errors and a class of power entropies. They numerically compared tightness of the relationship between the power entropies and the Bayes error and showed that the quadratic entropy seems to give at average the most accurate estimate of the posterior Bayes error.

We deal with the tightness of relationship between the Bayes errors and the family of f -entropies, introduced by Zvárová [7], which include many of the functions used for measuring biological diversity with applications in biology and medicine,

i.e. in the cancer research (Maley et al. [5]), or in the research of relations among human populations (Chak-raboty [6]). Some of the power entropies are also included in the f -entropies family. We propose an approximation of the lower bound of posterior Bayes error under given f -entropy that simplifies the formula for average inaccuracy. Then we show analytically that the quadratic entropy has the tightest relation to the prior Bayes error among f -entropies under our approximation.

2. The f -entropy

A generalization of entropy, called f -entropy, was suggested by Zvárová [7]. Here we will briefly recall its derivation.

Assume that $p(x)$ is absolutely continuous with respect to $q(x)$, i.e. $q(x) = 0 \Rightarrow p(x) = 0$, $x \in \chi$ and the convention $0f(0/0) = 0$ is used. The information divergence

$$D(p(x) \| q(x)) = \sum_{x \in X} p(x) \ln(p(x)/q(x))$$

of two discrete distributions $p(x)$ and $q(x)$ defined on the space χ was proposed by Kullback and Liebler [8].

The entropy

$$H(X) = - \sum_{x \in X} p(x) \ln p(x)$$

and the mutual information

$$I(X; Y) = \sum_{x \in X} \sum_{y \in Y} p(x, y) \ln \frac{p(x, y)}{p(x)p(y)}$$

were introduced by Shannon [9]. They satisfy equalities

$$I(X; Y) = D(p(x, y) \| p(x)p(y))$$

$$H(X) = I(X; X) \quad (1)$$

and inequality

$$0 \leq I(X; Y) \leq H(X)$$

A more general class of divergences, called f -divergences, was proposed by Csiszár [10]. The f -divergence of $q(x)$ from $p(x)$ is

$$D_f(p(x) \| q(x)) = \sum_{x \in X} q(x) f\left(\frac{p(x)}{q(x)}\right)$$

for any function f that is convex on the interval $[0, \infty)$, strictly convex in $f(1)$ and $f(1) = 0$. Following Zvárová [7], we can replace the information divergence in (1) by the f -divergence and define the generalized mutual information by

$$I_f(X; Y) = D_f(p(x, y) \| p(x)p(y)).$$

Let f be a function that satisfies the conditions required when defining f -divergences and also let $(f(x) - f(0))/x$ be a concave function on $[1, \infty)$. The f -entropy is then defined by the formula

$$H_f(X) = I_f(X, X) = \sum_{x \in X} p^2(x) f(1/p(x)) + f(0) \left(1 - \sum_{x \in X} p^2(x)\right).$$

Zvárová [7] showed that the f -entropies share many properties with the Shannon's entropy, e.g. that under given $r = \text{card}(\chi)$ they attain their minimal possible value (zero) when X is a constant with probability equal to one, attain their maximal possible value when X is uniformly distributed on χ and they satisfy the inequality

$$0 \leq I_f(X; Y) \leq H_f(X).$$

Further information about f -entropies, their characteristics and usage can be found in the paper of Zvárová, Vajda [11] and in the work Horáček [12].

3. The average inaccuracy

Let $r = \text{card}(\chi) \geq 2$ and denote $\Delta_r = \{(p(x_1), \dots, p(x_r)) : p(x_i) \geq 0, \sum_{i=1}^r p(x_i) = 1\}$. The bounds of f -entropy under given prior Bayes error are given by

$$H_f^+(e) = \max_{\{X: e(X)=e\}} H_f(X)$$

respectively by

$$H_f^-(e) = \min_{\{X: e(X)=e\}} H_f(X),$$

i.e. the maximum (minimum) is taken with respect to any discrete distribution $p(x)$ of random variable X on χ that satisfies $e(X) = 1 - \max_{x \in \chi} p(x) = e$. Since any f -entropy is a Schur-concave function on Δ_r (see e.g. Horáček [12]), the bounds satisfy

$$H_f^+(e) = H(1-e, e/(r-1), \dots, e/(r-1)) =$$

$$= (1-e)^2 f\left(\frac{1}{1-e}\right) + \frac{e^2}{r-1} f\left(\frac{r-1}{e}\right) + ef(0) \left(2 - e \frac{r}{r-1}\right)$$

when $e \in [0, (r-1)/r]$ and

$$H_f^-(e) = H(1-e, \dots, 1-e, 1-(1-e)k, 0, \dots, 0) =$$

$$= k(1-e)^2 f\left(\frac{1}{1-e}\right) + [1-k(1-e)^2] f\left(\frac{1}{1-k(1-e)}\right) +$$

$$+ f(0) [k(1-e)^2 + 1 - (1-k(1-e))^2]$$

when $e \in [(k-1)/k, k/(k+1)]$, $k \in \{1, \dots, r-1\}$.

Since the function $(f(x) - f(0))/x$ is concave on the interval $[1, \infty)$ and therefore also the function $x^2[f(x^{-1}) - f(0)]$ is concave on the interval $[0, 1]$, it is not difficult to show that the upper bound $H_f^+(e)$ is concave on the interval $[0, (r-1)/r]$ and the lower bound $H_f^-(e)$ is piecewise concave on each of the intervals $[(k-1)/k, k/(k+1)]$ where $k \in \{1, \dots, r-1\}$. Moreover, the convexity of f implies that the piecewise linear function

$$\Phi(t) = H_f\left(\frac{k-1}{k}\right) + \frac{t - \frac{k-1}{k}}{\frac{k}{k+1} - \frac{k-1}{k}} \cdot$$

$$\cdot \left[h_f\left(\frac{k}{k+1}\right) - h_f\left(\frac{k-1}{k}\right) \right]$$

when $t \in [(k-1)/k, k/(k+1)]$, $k \in \{1, \dots, r-1\}$ is convex on the interval $[0, (r-1)/r]$.

The upper bound for f -entropy under given posterior Bayes error e_B is defined by

$$H_f^+(e_B) = \max_{\{(X,Z): e_B(X|Z)=e_B\}} H_f(X|Z),$$

i.e. the maximum is taken with respect to any discrete random variable X defined on χ and random variable Y that satisfy $e_B(X|Z) = e_B$. The lower bound is defined accordingly. Since the bounds for f -entropy under prior Bayes error satisfy the concavity and convexity rules mentioned above, the proof of the theorem 4.1 in Morales and Vajda [1] can be followed to show that

$$\begin{aligned} H_f^+(e_B) &= (1-e_B)^2 f\left(\frac{1}{1-e_B}\right) + \\ &+ \frac{e_B^2}{r-1} f\left(\frac{r-1}{e_B}\right) + \\ &+ e_B f(0) \left(2 - e_B \frac{r}{r-1}\right) \end{aligned}$$

when $e \in [0, (r-1)/r]$ and that the lower bound is a convex, piecewise linear function with values

$$H_f^-(e_B) = (1-e_B) f\left(\frac{1}{1-e_B}\right) + f(0) e_B$$

when $e_B = \frac{k}{k+1}$, $k \in \{0, \dots, r-1\}$,

i.e. $H_f^-(e_B) = \Phi(e_B)$.

Since these bounds are continuous strictly increasing functions on the interval $[0, (r-1)/r]$, they implicitly set the bounds for the posterior Bayes error under given entropy. Following Morales and Vajda, we denote these bounds

$$\begin{aligned} e_{B,f}^+(H) &= \max_{H_f^-(e_B) \leq H} e_B \\ e_{B,f}^-(H) &= \max_{H_f^+(e_B) \leq H} e_B. \end{aligned}$$

Let us also denote

$$a_{r,f} = H_f\left(\frac{1}{r}, \dots, \frac{1}{r}\right)$$

and recall that the power entropies are defined by the formula

$$H_\alpha(X) = \frac{1}{\alpha-1} \left(1 - \sum_{x \in X} p^\alpha(x)\right)$$

when $\alpha > 0$.

(The entropy H_1 is defined as a limit and is equal to the Shannon's entropy.) Note that the power entropies are f -entropies when $\alpha \in (0, 2]$, namely the quadratic entropy $H_2(X)$ is f -entropy with $f(x) = x-1$.

Morales and Vajda introduced a measure of tightness of the relationship between the power entropies and the posterior Bayes error, called the average inaccuracy. It is defined by the formula

$$\begin{aligned} AI_r(e_B | H_\alpha) &= \frac{1}{a_{r,\alpha}} \cdot \\ &\cdot \int_0^{a_{r,\alpha}} [e_{B,\alpha}^+(H) - e_{B,\alpha}^-(H)] dH. \end{aligned}$$

This measure can be straightforwardly generalized to any strictly concave f -entropy by replacing $\alpha_{r,\alpha}$ by $\alpha_{r,f}$ and $e_{B,\alpha}$ by $e_{B,f}$ in the formula.

Morales and Vajda stated, on the basis of numerical computations with various values of the parameter α , that the quadratic entropy seems to have the lowest average inaccuracy when estimating posterior Bayes error among power entropies. However, due to the form of the lower bound for the posterior Bayes error, it is difficult to verify this postulate analytically. We suggest to approximate the piecewise linear function $H_f^-(e_B)$ by a function

$$\tilde{H}_f^-(e_B) = (1-e_B) f\left(\frac{1}{1-e_B}\right) + f(0) e_B$$

that acquire the same values as the original function when $e_B = k/(k+1)$, $k \in \{0, \dots, r-1\}$. This alternative lower bound is also a convex function on the interval $[0, (r-1)/r]$, it acquires the same values as the original piecewise linear bound in $\{(k-1)/k, k \in \{1, \dots, r-1\}\}$ and therefore it cannot be higher than the original bound. Since $H_f^-(e_B)$ is again a continuous strictly increasing function, the inverse function $\tilde{e}_{B,f}^-(H)$ is well defined and we can approximate the average inaccuracy by

$$\tilde{AI}_r(e_B | H_f) = \frac{1}{a_{r,f}} \int_0^{a_{r,f}} [\tilde{e}_{B,f}^+(H) - \tilde{e}_{B,f}^-(H)] dH.$$

The approximated bound is smooth when the function f is smooth. An example of the bounds $H^+(e_B)$, $H^-(e_B)$ and $\tilde{H}^-(e_B)$ is displayed in the figure 1.

Using this approximation, the following theorem can be formulated.

Theorem 1 Under the approximation described above, the quadratic entropy $H_2 = 1 - \sum_{x \in X} p^2(x)$ has the tightest relation to posterior Bayes error, i.e.

$$\tilde{AI}_r(e_B | H_2) \leq \tilde{AI}_r(e_B | H_f)$$

for any f -entropy H_f and any given integer $r \geq 2$.

Proof. Recall that when H_f is a f -entropy, the function f is convex on the interval $[0, \infty)$ and strictly convex in $f(1)$ with $f(1) = 0$. Let us first consider only those functions f that satisfy $f(0)=0$. Then it follows that $f(x)$ has to be above zero for all $x > 1$. Since $H_f^+(0) = H_f^-(0)$ and also $H_f^+((r-1)/r) = H_f^-((r-1)/r)$, we can calculate the approximate average inaccuracy $\tilde{AI}_r(e_B, H_f)$, i.e. the relative area between bounds $\tilde{e}_{B,f}^+(H)$ and $\tilde{e}_{B,f}^-(H)$, by calculating the area between bounds $H_f^+(e_B)$ and $H_f^-(e_B)$ then applying the corresponding normalizing constant.

$$\begin{aligned}
\tilde{A}I_r(e_B | H_f) &= \\
&= \frac{1}{a_{r,f}} \int_0^{a_{r,f}} [\tilde{e}_{B,f}^+(H) - \tilde{e}_{B,f}^-(H)] dH = \\
&= \frac{1}{a_{r,f}} \int_0^{\frac{r-1}{r}} [H_f^+(e_B) - \tilde{H}_f^-(e_B)] de_B = \\
&= \frac{r}{f(r)} \int_0^{\frac{r-1}{r}} (1-e_B)^2 f\left(\frac{1}{1-e_B}\right) + \\
&\quad + \frac{e_B^2}{r-1} f\left(\frac{r-1}{e_B}\right) - (1-e_B) \cdot \\
&\quad \cdot f\left(\frac{1}{1-e_B}\right) de_B = \frac{r}{f(r)} \int_0^{\frac{r-1}{r}} (1-e_B)^2 \cdot \\
&\quad \cdot f\left(\frac{1}{1-e_B}\right) + \frac{e_B^2}{r-1} f\left(\frac{r-1}{e_B}\right) - \\
&\quad - (1-e_B) f\left(\frac{1}{1-e_B}\right) de_B = \frac{r}{f(r)} \cdot \\
&\quad \cdot \int_0^{\frac{r-1}{r}} e_B \left[f\left(\frac{r-1}{e_B}\right) \frac{e_B}{r-1} - f\left(\frac{1}{1-e_B}\right) - \right. \\
&\quad \left. - (1-e_B) \right] de_B
\end{aligned}$$

After transformation $z = (r/r-1)e_B$ we have

$$\begin{aligned}
\tilde{A}I_r(e_B | H_f) &= \frac{r}{f(r)} \int_0^1 \left(\frac{r-1}{r}\right)^2 z \cdot \\
&\quad \cdot \left[f\left(\frac{r}{z}\right) \frac{z}{r} - f\left(\frac{r}{r-zr+z}\right) \frac{r-zr+z}{r} \right] dz
\end{aligned}$$

(2).

Since f is convex, $z \in (0,1]$ and $r \geq 1$, we get following inequality

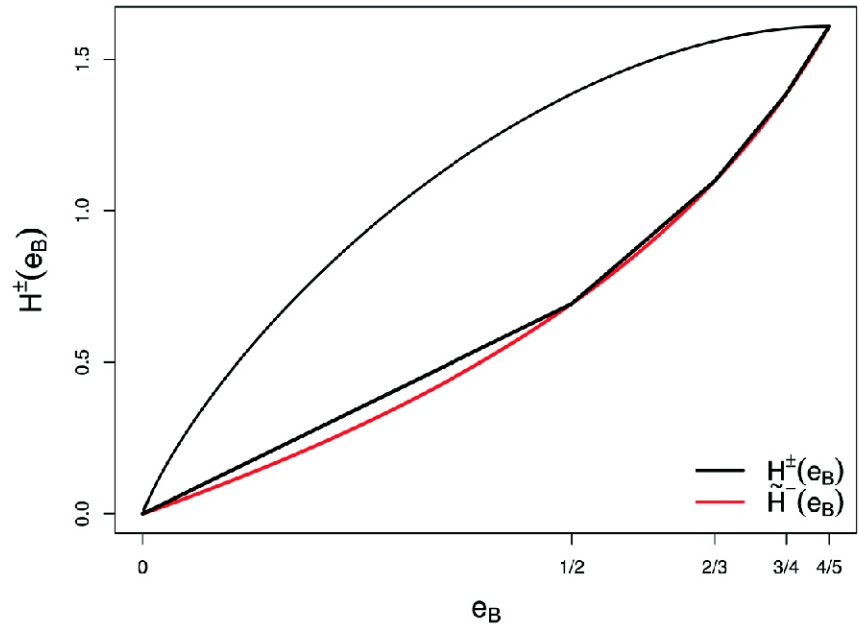


Fig. 1. The upper and lower bounds $H^\pm(e_B)$ and the approximate lower bound $\tilde{H}^-(e_B)$ for Shannon's entropy.

$$\frac{f(r/z) - f(1)}{r/z - 1} \geq \frac{f(r) - f(1)}{r - 1}$$

$$\frac{f(r/z)}{r/z - 1} \geq \frac{f(r)}{r - 1}$$

$$\frac{f(r/z)z}{f(r)} \geq \frac{r-z}{r-1}$$

(3).

Similarly the formula

$$\frac{f[r/(z-zr+r)] - f(1)}{r/(z-zr+r) - 1} \leq \frac{f(r) - f(1)}{r - 1}$$

can be converted into

$$\frac{f[r/(z-zr+r)](z-rz+r)}{f(r)} \leq z$$

(4).

After inserting inequalities (3) and (4) into the equality (2) we get

$$\begin{aligned}
\tilde{A}I_r(e_B | H_f) &\geq \int_0^1 \left(\frac{r-1}{r}\right)^2 z \left(\frac{r-z}{r-1} - z\right) dz \\
&= \int_0^1 \left(\frac{r-1}{r}\right) z(1-z) dz = \frac{1}{6} \frac{r-1}{r} = \tilde{A}I_r(e_B | H_2).
\end{aligned}$$

Now let us consider the situation when $f(0)=0$. Then $\tilde{A}I_r(e_B | H_f)$ satisfies

$$\begin{aligned}
\tilde{A}I_r(e_B | H_f) &= \\
&= \frac{1}{a_{r,f}} \int_0^1 \left(\frac{r-1}{r}\right)^2 z \left[f\left(\frac{r}{z}\right) \frac{z}{r} - f\left(\frac{r}{r-zr+z}\right) \frac{r-zr+z}{r} \right] dz + \\
&\quad + \frac{1}{a_{r,f}} f(0) \frac{r-1}{r} \tilde{A}I_r(e_B | H_2).
\end{aligned}$$

Here we use the inequalities (3) and (4) again (multiplied by $f(r)/r$ to avoid a potential division by zero or by a negative number) and we receive

$$\tilde{A}I_r(e_B | H_f) \geq \frac{1}{a_{r,f}}.$$

$$\cdot \left[\left(\frac{f(r)}{r} + f(0) \frac{r-1}{r} \right) \tilde{A}I_r(e_B | H_2) \right]$$

$$\tilde{A}I_r(e_B | H_f) = \frac{r}{f(r) + (r-1)f(0)}.$$

$$\cdot \left[\frac{f(r) + f(0)(r-1)}{r} \tilde{A}I_r(e_B | H_2) \right] =$$

$$= \tilde{A}I_r(e_B | H_2).$$

4. Conclusions

This result covers the power entropies H_α when $\alpha \in (0,2]$ and also other f -entropies, for example Emlen's index [13]

$$H_E(p(x)) = \sum_x p(x) e^{-p(x)} - e^{-1},$$

where

$$f(x) = x \left(e^{-1/x} - e^{-1} \right) \chi_{(0,\infty)}(x),$$

Good's index [14]

$$H_{G(\alpha,\beta)}(p(x)) = \sum_x p(x)^\alpha (-\ln p(x))^\beta$$

when $\alpha \in [1/2, 1]$, $\beta = 1$, where

$$f(x) = x^{2-\alpha} (\ln x)^\beta I_{(1,\infty)}(x)$$

its generalization, the index of Sharma and Taneja [15]

$$H_{ST(s,t)}(p(x)) = (s-t)^{-1} \sum_{i \in M} (p(x)^t - p(x)^s)$$

when $t \in [0,1]$, $s \in (\max\{t, 1-t\}, 3-t]$ (defined as a limit when $s=t$) for

$$f(x) = (s-t)^{-1} (x^{2-t} - x^{2-s}) \chi_{[1,\infty)}(x),$$

and the index of Ferreri [16]

$$H_{Fe(\alpha)}(p(x)) = (1+\alpha)\alpha^{-1} \ln(1+\alpha) -$$

$$-\alpha^{-1} \sum_x (1+\alpha p(x)) \ln(1+\alpha p(x))$$

$$\alpha > 0$$

where

$$f(x) = \left[\frac{x(1+\alpha)}{\alpha} \ln(1+\alpha) - \right.$$

$$\left. - \frac{x(x+\alpha)}{\alpha} \ln \frac{x+\alpha}{x} \right] I_{(0,\infty)}(x).$$

Acknowledgment

This work was supported by the project 1M06014 MŠMT ČR and SVV-2010-265513. The author thanks I. Vajda for introduction to the problem, and J. Zvárová and M. Studený for various useful suggestions during the preparation of the work.

References

- [1] Morales D., Vajda I.: Generalized Information Criteria for Optimal Bayes Decisions. Research report 2274. ÚTIA AV ČR, v.v.i.; 2010.
- [2] Fano R.M.: Transmission of Information: A Statistical Theory of Communications. New York: MIT Press; 1961.
- [3] Kovalovsky V.A.: The problem of character recognition from the point of view of mathematical statistics. Character Readers and Pattern Recognition. New York: Spartan Books; 1968;3-30.
- [4] Vajda I., Vašek K.: Majorization, concave entropies and comparison of experiments. Problems of Control and Information Theory. 1985;14:105-15.
- [5] Maley C.C., Galipeau P.C., Finley J.C., Wongsurawat V.J., Li X., Sanchez C.A., Paulson T.G., Blount P.L., Risques R.A., Rabinovitch P.S., Reid B.J.: Ecological Diversity Measures Predict Cancer Progression In Barrett's Esophagus. Gastroenterology. 2006;131:672-673.
- [6] Chakraborty R.: Mitochondrial DNA Polymorphism Reveals Hidden Heterogeneity within Some Asian Populations. American Journal of Human Genetic. 1990;47:87-94.
- [7] Zvárová J.: On Measures of Statistical Dependence. Časopis pro pěstování

matematiky. 1974;99:15-29.

- [8] Kullback S., Leibler R.A.: On Information and Sufficiency. Annals of Mathematical Statistics. 1951;22:7986.
- [9] Shannon C.E.: A mathematical theory of communication. Bell System Technical Journal. 1948;27:379-423,623-56.
- [10] Csizsár I.: Eine informationstheoretische Ungleichung und ihre Anwendung auf den Beweis der Ergodizität von Markoffschen Ketten. Publ. Math. Inst. Hungar. Acad. Sci. 1963;8:85-107.
- [11] Zvárová J., Vajda I.: On Genetic Information, Diversity and Distance. Methods of Inform. in Medicine. 2006;2:173-179.
- [12] Horáček M.: Míry biodiverzity a jejich aplikace. (Measures of biodiversity and their applications.) Master thesis. Prague, Charles university. Supervisor J. Zvárová. 2009.
- [13] Emlen J.M.: Ecology: An Evolutionary Approach. Reading: Addison-Wesley; 1973.
- [14] Good I.J.: The population frequencies of species and the estimation of population parameters. Biometrika. 1953;40:237-64.
- [15] Sharma B.D., Taneja I.J.: Entropy of type (;) and other generalized measures in information theory. Metrika. 1975;22:205-15.
- [16] Ferreri C.: Hypoentropy and related heterogeneity divergence measures. Statistica. 1980;40:55-118.

Contact

Mgr. Martin Horáček

Centre of Biomedical Informatics,
Institute of Computer Science,
Academy of Sciences of the Czech
Republic
Pod Vodárenskou věží 2
182 07 Prague 8
Czech Republic
e-mail: horacek@euromise.cz

Approaches for Constructing Age-Related Reference Intervals and Centile Charts for Fetal Size

M. Hynek^{1,2}

¹Gennet, Center for Fetal Medicine, Prague, Czech Republic

²Department of Gynecology and Obstetrics, Thomayer University Hospital, Prague, Czech Republic

Supervisor: Prof. RNDr. Jana Zvárová, DrSc.

Summary

The assessment of the fetal size based on ultrasound biometry is important for proper pregnancy management. The means for evaluations of these measurements represent age-related reference intervals and centile charts which are used to detect the extreme values, possibly indicating pathology. The key features concerning the design of studies used for the construction of these intervals are given. The aim of this paper is to review possible statistical approaches, pointing out the methodology, goodness of fit, advantages and limitations.

Keywords: Fetus, gestational age, ultrasound, biometry, pathology, reference interval, centile, centile chart, normal distribution, normalizing transformation, quantile regression

1. Introduction

The assessment of fetal size is of a crucial importance for the proper pregnancy management. It is implicitly based on ultrasound measurements of embryonic and fetal biometrical parameters (e.g. crown-rump length, head circumference, femur length, etc.), most of which almost invariably increase with the gestational age (GA) [1]. The means for evaluation of these measurements are age-related reference intervals (RIs) and centile charts, allowing interpretation of obtained fetal body measurement in comparison with the body size of fetuses in the reference population [2].

The RI (often misleadingly called 'normal range') represents the interval between a pair of predetermined extreme centiles (commonly the 5th and 95th for a 90% interval) of a size variable, denoted y , at a given GA (denoted t). Centile charts plot the value of y corresponding to one or more

centiles against relevant GA over a range of GAs.

In the field of fetal size, measured values of y which lie outside the RI are regarded as extreme and may indicate the presence of underlying pathology (e.g. morphological or chromosomal abnormality [3], intrauterine growth restriction [4], macrosomia [5]). Otherwise, more informatively, the value's exact centile position relative to the reference population is estimated from the knowledge of the distribution of y at a given GA. The proximity of the centile position to 0% or 100% is a measure of how extreme the observation is compared to the reference population [6].

Variety of strategies for constructing of RIs and centile charts has been published. The choice of appropriate methodology in the field of fetal biometry is especially crucial as inaccurate centiles may lead to incorrect conclusions regarding fetal development. The aim of this article is to scrutinize possible statistical methods for age-related RIs and centile charts for fetal size.

2. Study design

Prior to statistical methods we consider the key features of the study design.

Size and growth. It is important to bear in mind the difference between two notions, which are commonly confused: size and growth. The fetal size represents a single measurement point, while fetal growth refers to a dynamic process [7], [8]. These two concepts contain completely diverse information. The fetal size can be normal despite abnormal growth, whereas normal fetal growth can be associated with an abnormal size. It is necessary to retain the clear distinction between the size and

growth and to realize that the appropriate study design for these outcomes is different.

Cross-sectional and longitudinal studies. The same as above applies for the difference between cross-sectional and longitudinal studies [1], [4]. In the former, each fetus contributes just once to a reference sample, whereas in the latter each fetus is measured several times. Therefore, cross-sectional data give information only on the size and can be used to compare the size of a fetus (of known GA) on a single occasion with reference data.

Longitudinal data may be used to produce RIs for the fetal size as well as growth. The analysis of longitudinal data requires a different approach. Serial measurements on an individual fetus are highly correlated so that the effective sample size in such an approach is likely to be nearer to the number of fetuses than to the total number of observations [1]. Thus, statistical methodology must be accordingly adjusted. Further information can be found in the literature [9], [10]. The following review will deal with the constructing RIs for the fetal size based on the cross-sectional data, as they are more common in the field of fetal biometry, maybe because of easier data collection compared to longitudinal studies.

Sample selection and data collection.

The choice of an appropriate sample is of a great importance. It is preferable to collect data specifically for the purpose of developing RIs, with each fetus being included only once [1]. The aim is to obtain as unselected and representative group as possible with no prior selection for examination as data collected at a scan done for any clinical indication may seriously bias the results.

Typically, the status of pregnancy is determined retrospectively, and there is scope for controversy in deciding which pregnancy should be regarded as abnormal and excluded from the reference sample. Altman and Chitty [1] recommend using for the exclusion only information available at the time of the ultrasound measurement, with the exceptions of subsequently found serious congenital anomaly or condition affecting fetal growth, such as maternal diabetes or renal disease. There is desirable to have approximately equal numbers of measurements at each week of gestation [1].

Sample size. It is hard to specify the appropriate sample size for developing centile charts. The larger the sample size the greater precision centiles will have. The interest is usually focused on the tails of the distribution. Royston [11] published that the standard error (SE) of the 100 α th centile ($c_{100\alpha}$) can be approximately expressed as a multiple of the age-specific standard deviation (SD) as

$$SE(c_{100\alpha}) \approx SD \sqrt{\frac{1 + \frac{1}{2} z_{1-\alpha/2}^2}{n}} \quad (1a),$$

where $z_{1-\alpha/2}$ is the appropriate value from the standard normal distribution and n is the number of cases [11]. Therefore for the sample size n we obtain

$$n \approx \frac{1 + \frac{1}{2} z_{1-\alpha/2}^2}{\left(\frac{SE(c_{100\alpha})}{SD}\right)^2} \quad (1b).$$

If, for example, we require SE for a 90% reference interval to be 5% of SD, the resulting value of n would be $n \approx (1 + 0.5 \cdot 1.645^2)/0.05^2 \approx 941$, for SE corresponding to 10% of SD $n \approx 235$. Besides, more extreme centiles are less precisely estimated. It is clear that basically several hundred observations are necessary to get reasonable estimates of extreme centiles.

3. Statistical methods

The number of possible statistical methods for the calculation of RIs and centile charts has mushroomed in the last two decades.

The basic requirements for the method are to produce centiles which change smoothly with GA and provide good fit to the data [1]. The requirement stated in the nineties, to maintain as simple as possible [1], is becoming less important due to the broad availability of a computer technology. Finally, it is desirable that model allows the calculation of the relevant centile position and Z-score for any further measurement [12]. Z-scores have been used increasingly in recent years, and are the WHO-recommended system for comparing individual anthropometric measurement with the reference population [13]. They have also proved to be a powerful quality-control tool, allowing the sonographers to choose appropriate reference charts [14] and to audit their daily practice [15].

Generally, methods are divided into two broad categories: parametric (based on modelling the distribution) and non-parametric (empirical).

3.1 Mean and SD model

In this most common parametric method, the basic assumption is that at each GA the measurement of interest has a normal distribution with a mean and SD that vary smoothly with GA. A desired centile curve is then calculated using formula

$$c_{100\alpha} = \mu + k \cdot SD \quad (2),$$

where k is the corresponding centile of the normal distribution (e.g. for 10th and 90th centile k is ± 1.28 , for 5th and 95th centile k is ± 1.645 , etc.) and μ and SD are, respectively, the mean and standard deviation at the required GA for the reference population. The approach is based upon least-squares regression analysis to model both the mean and the SD curves as polynomial functions of GA [1], [4], [11], [16].

First, the regression analysis is applied to the raw data to find a suitable curve for the mean (Figure 1). As for the choosing the degree of polynomial, Royston [11] and Royston and Wright [4] recommend the initial use of a cubic polynomial ($a + b.t + c.t^2 + d.t^3$). If the coefficient d is not significantly different from zero (d divided

by its SE and the result referred to the appropriate point of the Student t distribution with $n - 4$ degrees of freedom), a quadratic polynomial is then fitted and the same assessment applied for c coefficient. The process is repeated until no further removal of terms is possible. Altman and Chitty [7] suggest the linear-cubic model ($a + b.t + c.t^3$) as a good alternative for fetal size data. Fitting polynomial of a higher degree than cubic is not advocated as the fitted curve may exhibit unrealistic features such as waviness or sharp deviation at extreme ages [4]. In case coefficients for quartic or quintic polynomials are statistically significant, the use of fractional polynomials is recommended (see section 3.1.2) [17]. The choice of curve should be based not only on statistical significance, but also the aesthetical appearance and the quality of fit to the data should be taken into account.

Next step is to check the goodness of fit of the obtained polynomial by plotting the residuals (observed minus predicted values) against GA to assess if and how variability changes with age [7]. The normal quantile-quantile (Q-Q) plot of residuals can reveal any departure from normality (e.g. positive or negative skewness). In this case transformation of the data towards normality should be tried prior to the analysis. See section 3.1.1.

After the suitable mean model is established, the attention is turned to the modelling of the variability. Many previously published studies fit the model which does not allow the SD to change with gestation. However, in the field of fetal measurements the between-subject variability almost always shows an upward trend during pregnancy [7], so it is crucially important to consider the relation between the SD and age. The approach of Altman [16] is one of the most frequently used. Based on the assumption that variable of interest has a normal distribution at all ages, the residuals from the mean model should be then also normally distributed. Subsequently, the absolute values of residuals should have a half normal distribution.

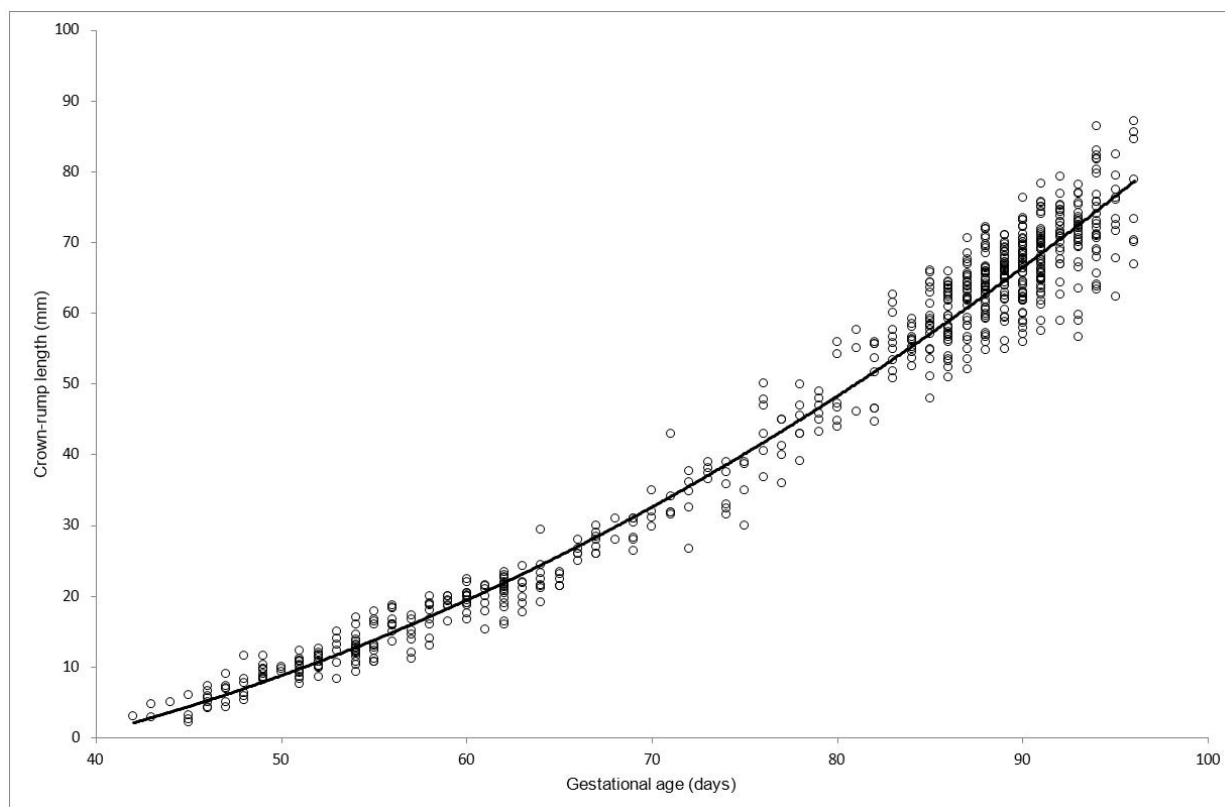


Fig. 1. Plot illustrating the relationship between fetal crown-rump length and gestational age during first trimester of pregnancy, showing the raw data and the fitted curve for the mean (—) with the quadratic polynomial $y = 0.0126 t^2 - 0.3177 t - 6.7365$. The sample consists of 664 fetuses with exactly known gestational age (in-vitro fertilization pregnancies).

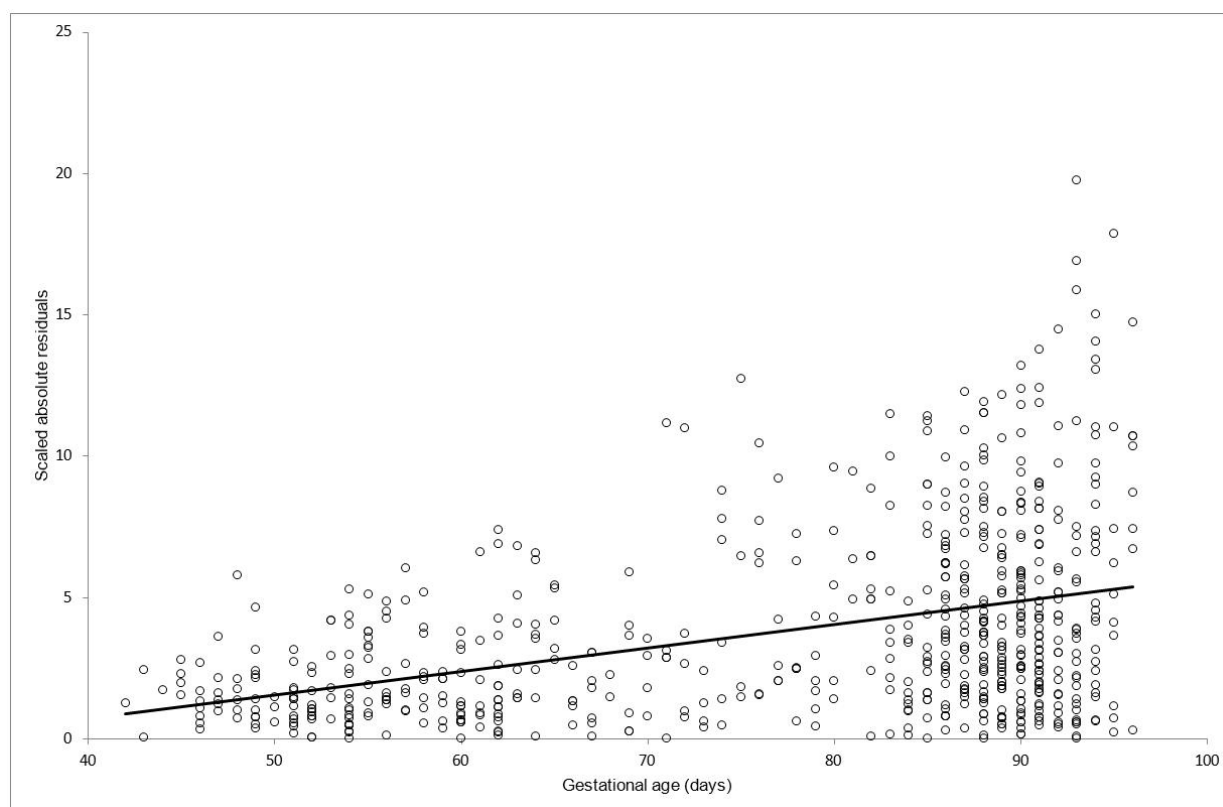


Fig. 2. Plot of scaled absolute residuals from the regression line shown in Figure 1 together with fitted linear regression line $y = 0.0833 t - 2.6173$.

If we have a variable X which follows a normal distribution with mean zero and variance 2, the absolute value $|X|$ follows a half-normal distribution which has mean $\sqrt{(2/\pi)}\sigma$ [18]. The mean of the absolute residuals multiplied by $\sqrt{(\pi/2)}$ is an estimate of the SD of the residuals. Thus if the SD is not reasonably constant over age, the predicted values from regression of the absolute residuals on age multiplied by $\sqrt{(\pi/2)}$ will give age-specific estimates of the SD of the residuals, and hence of y .

An identical approach published by Royston and Wright [4] comprises to produce 'scaled absolute residuals' (SARs) by multiplying the absolute residuals by $\sqrt{(\pi/2)}$. Subsequently, SARs are regressed on age in the same way as for the mean and again the predicted values from this curve estimate the SD of the residuals (Figure 2). It is unlikely that a curve more complex than quadratic is needed for a satisfactory fit for the SD [16]. Superimposing 1.645 SD on the residual plot is helpful in assessing how well the SD

has been modelled, as presumably 90% of the observed residuals should fall within these limits.

If absolute residuals show no trend with age, the SD is estimated as the SD of unscaled original residuals.

The regression analysis to estimate the mean ought strictly to take into account any change in SD with gestation. Thus the mean model could be subsequently refitted using the reciprocal of the square of the estimated age-specific SD as weights [19]. However, the effect of this refitting is almost always rather small [7].

The next very important step includes the evaluation of goodness of fit. A useful tool includes Z-scores which we calculate for the observed values y using formula

$$Z \text{ skór} = \frac{y - \mu_{GS}}{SD_{GS}} \quad (3),$$

where GA and SDGA are, respectively, the mean and SD given by the model for the GA at which the observation y is made. Then several modalities are advocated [7]: a plot of Z-scores against age to confirm that no untoward pattern is present and to check that the expected proportions of values fall between or outside the appropriate centiles (for example 90% of Z-scores should lie between $Z = 1.645$, 5% above and 5% below) (Figure 3). Further normal Q-Q plot of Z-scores can be used to check if they have close to normal distribution (Figure 4). This is signified by roughly straight line and can be formally confirmed using Shapiro-Wilk W [20] or Shapiro-Francia W' test [21]. However, in large samples slight deviation from normality may lead to statistically significant non-normality, so it is unwise to worry about moderately small p (such as $p = 0.01$) unless the normal plot shows clear deviation from a straight line [7].

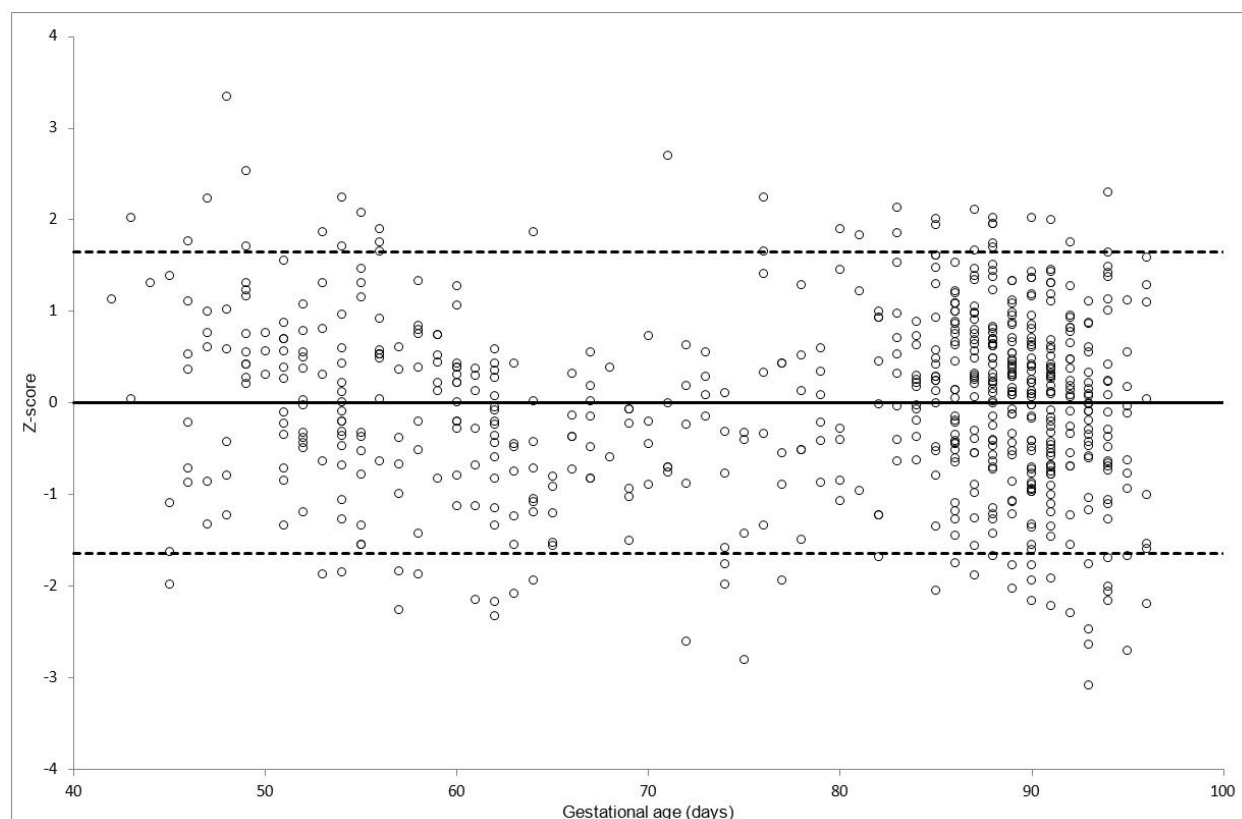


Fig. 3. Plot of Z-scores of observations with the expected 5th and 95th centiles 6.0% of observations fall below 5th centile, 5.3% above 95th centile and 88.7% between 5th and 95th centiles.

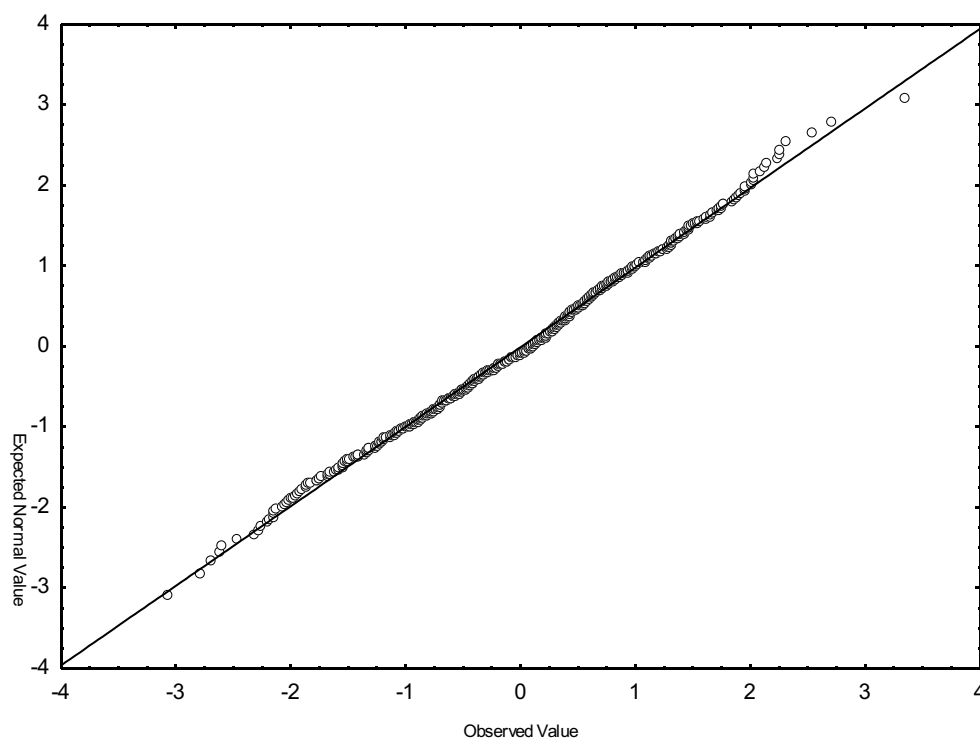


Fig. 4. Normal quantile-quantile plot of Z-scores of observations. Shapiro-Wilk W test = 0.997 ($p = 0.31$).

Eventually, a satisfactory model being determined, the centile curves for the desired RIs may be calculated by substituting the expression for the fitted

mean and SD into equation (2) and superimposed on a scatter diagram of observations as a final check to fit (Figure 5). The Z-score for any new measurement

can be calculated using equation (3) and its centile obtained from the inverse normal distribution function.

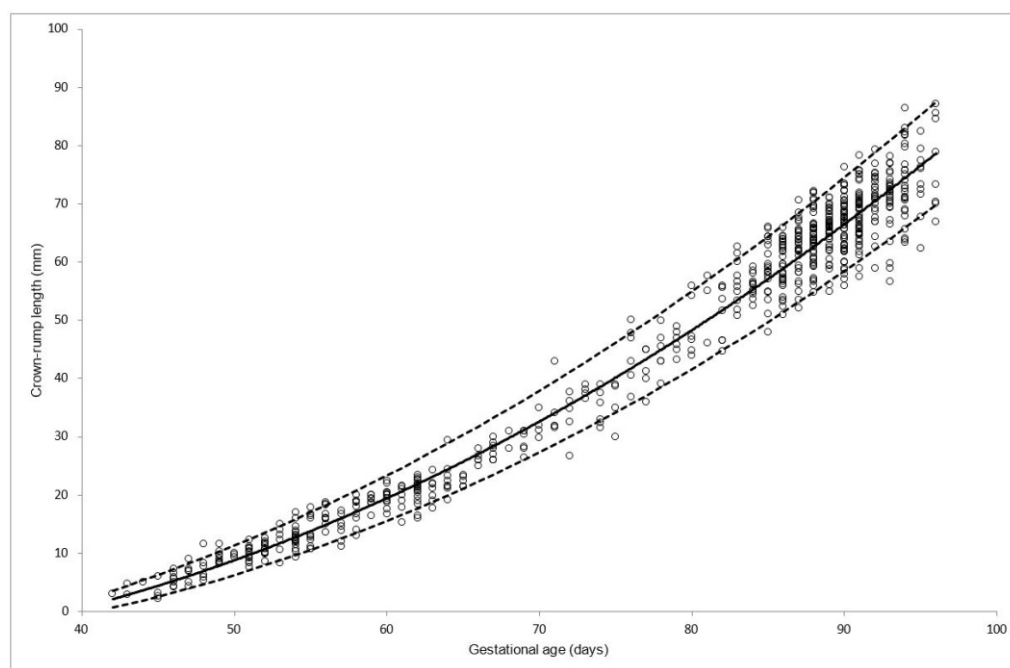


Fig. 5. Final model of fetal crown-rump length in relation to gestational age, showing raw data together with reference curves for the mean (—), 5th and 95th centiles (-----).

3.1.1 Logarithmic transformation

The method just mentioned requires the assumption that at each GA the data come from a population with normal distribution. However, this is not always the case, as many fetal measurements often present skewness in the distribution (mostly positive, with the right tail of the distribution longer than the left). The logarithmic transformation can help to overcome the issue. The solution at the same time stabilizes the variance in case the SD increases rapidly with age [11].

Royston [11] suggests if residuals from the initial model show a positive skew to perform a logarithmic transformation on the original value y and refit the model on $\log(y)$. If residuals from refitted model are once again skewed, then it is recommended to try a shifted logarithmic transformation of the form $\log(y + C)$, with $C > 0$ for residuals negatively skewed and $C < 0$ for positively skewed. The best value of C is that which maximizes p -value for the normality test of the residuals.

Once the model is finalized it is important to back-transform the curves using antilog (and to subtract C , respectively).

3.1.2 Fractional polynomials

Conventional polynomials suffer from several well-known disadvantages. Low order polynomials offer only a few curve shapes and thus do not always fit the data well, whereas high order ones may fit badly at the extremes of the observed range. Further, they do not have asymptotes and cannot fit data where limiting behaviour is expected [22]. Royston and Altman [17] proposed an extended family of curves, called fractional polynomials (FPs), whose power terms are restricted to a small predefined set of integer and non-integer values. Conventional polynomials are a subset of this family. If conventional polynomial is of a form

$$a + b.t + c.t^2 + d.t^3 + \dots \quad (4),$$

FPs are defined as

$$a + b.t^{p_1} + c.t^{p_2} + d.t^{p_3} + \dots \quad (5),$$

where p_1, p_2, \dots are chosen from the set $\{-2, -1, -0.5, 0, 0.5, 1, 2, 3\}$. Conventionally, the power 0 represents natural logarithmic transformation, so that t^0 equals to $\log_e(t)$. Another extension involves 'repeated powers'. In this case the second term is multiplied by $\log_e(t)$. For example, an FP of degree 3 with powers (0, 2, 2) is then of a form

$$a + b.\log_e(t) + c.t^2 + d.t^2.\log_e(t) \quad (6).$$

An FP of first degree is of the form $a + b.t^p$. For a given data set, the best value of p is found by fitting eight separate linear regressions using t^2, t^1, \dots, t^p and selecting the value of p which gives the best fit. Regarding FPs of the second degree ($a + b.t^{p_1} + c.t^{p_2}$), using the standard set detailed above would involve fitting a model for each of 36 permissible combinations of powers. The model with the lowest residual SD is chosen as the best.

The use of FPs can often give a better fit even with fewer terms compared to conventional polynomials.

3.2 LMS method

The LMS method, introduced by Cole [23], [24] and further refined by Cole and Green [25], is an extremely flexible and widely applicable semi-parametric method which can produce smooth centile curves even when the data appear to have a complex shape. Furthermore, time-varying skewness, which cannot be taken into account with classical logarithmic transformation, is easily dealt with. The method assumes that the use of suitable power transformation can remove skewness and normalize the data. Such a family of transformations is that proposed by Box and Cox [26], with the optimal power λ at a given age calculated from the data to completely remove the skewness.

The distribution of the variable of interest y changes smoothly with age and is completely summarized by three parameters λ (Box-Cox power, related to skewness), μ (median) and σ (coefficient of variation), the initials of which (L, M and S) give the name to the method.

Suppose that y has median μ and that y^λ (or $\log_e(y)$ if $\lambda = 0$) is normally distributed. Then we consider transformed variable x

$$x = \begin{cases} \left(\frac{y}{\mu}\right)^\lambda - 1 & \text{pro } \lambda \neq 0 \\ \log\left(\frac{y}{\mu}\right) & \text{pro } \lambda = 0 \end{cases} \quad (7)$$

based on the Box-Cox transformation [26]. This transformation maps the median of y to $x = 0$ and is continuous at $\lambda = 0$. For $\lambda = 1$ the SD of x is exactly the coefficient of variation (CV) of y , and this remains approximately true for all close to one [25]. The optimal value of λ is that minimizes the SD of x .

Denoting the SD of x (and the CV of y) by σ , the Z-score of x (and hence y) is given by

$$Z = \frac{x}{\sigma} = \begin{cases} \frac{\left(\frac{y}{\mu}\right)^\lambda - 1}{\lambda\sigma} & \text{pro } \lambda \neq 0 \\ \frac{\log\left(\frac{y}{\mu}\right)}{\sigma} & \text{pro } \lambda = 0 \end{cases} \quad (8)$$

and is assumed to take a standard normal distribution. Assume that the distribution of y varies with t (GA), and that λ , μ and σ at t are read off the smooth curve $L(t)$, $M(t)$ and $S(t)$. Then

$$Z = \begin{cases} \frac{\left(\frac{y}{M(t)}\right)^{L(t)} - 1}{L(t)S(t)} & \text{pro } L(t) \neq 0 \\ \frac{\log\left(\frac{y}{M(t)}\right)}{S(t)} & \text{pro } L(t) = 0 \end{cases} \quad (9).$$

Rearranging (9) shows that centile $c_{100\alpha}$ of y at t is given by

$$c_{100\alpha}(t) = \begin{cases} M(t)[1 + L(t)S(t)z_\alpha]^{1/L(t)} \\ M(t)e^{S(t)z_\alpha} \end{cases} \quad \text{pro } L(t) \neq 0$$

$$\text{pro } L(t) = 0 \quad (10),$$

where z_α is the normal equivalent deviate of size α . This shows that if L , M and S curves are smooth, then so are the centile curves. The equation (10) allows us to calculate and draw any desired centile curve and the equation (9) allows converting any individual measurement into Z-score.

The original method [23], [24] involved splitting the data into age groups, as narrow as possible, but at the same time with adequate numbers of measurements (at least 100 for the best result), then estimating L , M and S for each group using maximum likelihood and finally constructing a smooth curve by the means of cubic splines, kernel methods, polynomials or simply drawing by eye. The choice of age cut-offs between groups is arbitrary and thus could influence the final result.

Cole and Green [25] added a non-parametric aspect to the original LMS method by using penalized likelihood function. The curves $L(t)$, $M(t)$ and $S(t)$ are estimated by maximizing the penalized likelihood

$$\begin{aligned} \ell - \frac{1}{2}\alpha_\lambda \int \{L''(t)\}^2 dt - \\ - \frac{1}{2}\alpha_\mu \int \{M''(t)\}^2 dt - \\ - \frac{1}{2}\alpha_\sigma \int \{S''(t)\}^2 dt \end{aligned} \quad (11),$$

where α_λ , α_μ and α_σ are smoothing parameters and log-likelihood function ℓ is given by

$$\ell = \ell(L, M, S) =$$

$$= \sum_{i=1}^n \left(L(t_i) \log \frac{y}{M(t_i)} - \log S(t_i) - \frac{1}{2} z_i^2 \right) \quad (12).$$

Three integrals provide roughness penalties for the curves $L(t)$, $M(t)$ and $S(t)$, leading to natural cubic splines with knots at each distinct value of t . The advantage of this approach is that the subjective grouping step is removed, the entire dataset is treated as a single entity and the curve fitting across the age is controlled directly by smoothing parameters. No age cut-offs need to be specified, and L , M and S values at each age are used in turn to calculate the other two parameters. The only arbitrariness is the choice of α_λ , α_μ and α_σ or in practice equivalent degrees of freedom (EDF), calculated for each fitted curve as a function of these smoothing parameters, which give a more usable measure of the extent of the smoothing. The EDF of each L , M and S curves is a measure of complexity (e.g. EDF = 1 means a constant, EDF = 2 corresponds to the straight line, EDF = 3 almost a quadratic curve, and EDF ≥ 4 indicates progressively more complex curve shapes). The choice must strike balance between the fidelity to the data and the smoothness. Low values of EDF may lead to the oversmoothed curves whereas high values to the undersmoothed ones. The disadvantage is that the choosing of EDFs is somewhat subjective. However, as a very rough guideline, we may use a procedure for assessing goodness of fit by comparing the difference in deviance ($-2\log(\text{penalized likelihood})$) between two models where the total number of EDFs differed by e to a χ_e^2 -distribution [6]. The other procedure, which helps in finding proper smoothing values for EDFs, represents a worm plot, a very sensitive diagnostic tool proposed by Buuren and Frederiks [27]. The worm plot consists of a collection of detrended Q-Q plots of residuals (detrended means that each empirical quantile is subtracted from its corresponding unit normal quantile

[28]), which are split according to age. A model that fits data well is characterized by a flat worm-like string. A particular change in shape and location of worm corresponds to the type of misfit (e.g. in mean, variance, skewness, kurtosis).

3.3 LMSP method

The assumption of normality following the Box-Cox transformation may be violated by the presence of kurtosis, for which the transformation does not adjust.

The LMSP method of Rigby and Stasinopoulos [29] can be understood as a generalization of the LMS approach. It uses a Box-Cox power exponential distribution to try to overcome the issue of kurtosis. The distribution has four parameters denoted as μ , σ , ν and τ and , which may be interpreted as relating to location (median), scale (approximately CV), skewness (transformation to symmetry) and kurtosis- (power exponential parameter), respectively. The advantage is that this distribution provides a flexible model for both skewness and kurtosis (allowing either for platykurtosis or leptokurtosis).

The model is fitted by maximizing a penalized likelihood. Centile estimation proceeds in a manner not dissimilar to that of the LMS method.

3.4 HRY method

Healy, Rabash and Young [30] (hence HRY) proposed a non-parametrical procedure based on the technique of Cleveland [31] for a smoothing a scatter plot by using locally weighted regression. This approach makes no assumption about the nature of the distribution of measurements at a given age and centiles are estimated empirically. At the same time it is expected that both centiles themselves and the intervals between centile at a fixed age should behave smoothly. This is done by constrain that the spacings between centiles can be expressed as a low-order polynomial in the underlying Z-score.

The first stage involves obtaining the selected raw centiles. The data are sorted into ascending age order. The first k measurements are then selected, where k is a fraction of the total number available, typically 5-10%. A regression of these k measurements on age is fitted, the desired centiles are obtained from the ranked residuals, using interpolation when necessary and their values are plotted against the median age value of k points. This procedure has used points 1 to k . It is repeated successively using points 2 to $k+1$, 3 to $k+3$, ... until the whole span of ages has been covered.

The initial centile curves will be very irregular, so the second stage is needed to smooth them. A smooth curve for the i th centile can be estimated by the polynomial of degree p

$$y_i = a_{0i} + a_{1i}t + a_{2i}t^2 + \dots + a_{pi}t^p \quad (13).$$

The coefficients a_{ji} for a fixed j are then modelled as a polynomial in z_i , where z_i is the normal equivalent deviate of the i th centile, so that

$$a_{ji} = b_{j0} + b_{j1}z_i + b_{j2}z_i^2 + \dots + b_{jq_j}z_i^{q_j} \quad (14),$$

where the degree q_j of the polynomial may differ from one value of j to another. This restricts the distance between centiles and prevents the resulting curve to cross. Combining equations (13) and (14) gives a linear model which simultaneously fits all the centiles estimated in stage one, and this can be fitted by least squares.

The method allows a considerable amount of flexibility. However, in practice a good deal of experimentation may be needed to choose good values for the adjustable parameters k , p and q_j . Authors recommend to judge the suitability of a particular set by counting the points falling between adjacent centiles and comparing these counts with expected values. In theory, for any observation a corresponding Z-score can be calculated by solving a polynomial equation, though the order of the polynomial may realistically prohibit this [32].

As certain variables observed over a wide age range may need more complex curve shapes than a single polynomial could provide. Pan et al. [33] suggested dividing the data into contiguous age groups, fitting polynomials within each group and smoothing points where these meet by using an extra polynomial term. As an alternative, Goldstein and Pan [34] proposed that these age groups be defined initially and that the fitted response be made smooth across the join points by equating derivatives of the curves on either side.

3.5 Non-parametric quantile regression

Other approaches for constructing the reference curve include a range of methods based on non-parametric quantile regression [35]. These methods allow quantiles to be estimated as a smooth function of covariates without imposing parametric distributional assumption, thus they are valuable in case that any transformation method is not able to achieve normality over the full range of relevant ages. The approach is robust to the presence of outliers and moreover, the procedure is unsupervised since all the smoothing parameters are determined adaptively.

Gannoun et al. [36] proposed three methods using kernel estimation, local constant kernel estimation and double kernel estimation of conditional quantile curves. Regarding the choice of kernels, authors use the univariate standard normal and uniform densities and also give some guideline for the selection of the bandwidths. Another example of using quantile regression presented Wei et al. [37]. For more details we refer to the original papers [36], [37].

Regarding the goodness of fit, an elegant way presents the worm plot, as shown by Buuren [38].

4. Discussion

For constructing age-related quantiles there are several different methods available, each of them having both advantages and limitations. Hence, it is unlikely that a single one would be appropriate in all circumstances.

The parametric 'mean and SD model' benefits from being relatively simple and easy to use, with the necessary techniques available in most basic statistical packages. It must be emphasized that the validity of reference ranges depends critically on the assumption of normality. The approach is able to cope with some heteroscedasticity by modelling the SD as age-varying and skewed data may sometimes be corrected by logarithmic or shifted logarithmic transformation. However, time-varying skewness and non-normal kurtosis cannot be easily accommodated. The resulting centile curves and Z-scores can be expressed as explicit formulae. The fact that the method suffers from the well-known limitations of a polynomial curve shape can be greatly improved by using the family of fractional polynomials.

In the field of fetal measurements this is the approach which has been repeatedly used so far as a wide range of fetal biometric measurements available from ultrasound scanning fulfil sufficiently the assumption of normality. Nevertheless, it is not always tenable and alternative techniques are required.

The LMS method with penalized likelihood is extremely flexible and widely applicable producing convincing centile curves irrespective of the complexity of the curve shape. It can easily cope with time-varying skewness, though some non-normal kurtosis may remain. Appealing by-product of the method is that $L(t)$, $M(t)$ and $S(t)$ curves may be of interest in their own as they facilitate further investigation in the underlying structure of the dataset. Although succinct formulae for centile curves are unobtainable, Z-score for any individual measurement can be easily calculated. There are, however, some drawbacks with the smoothing approach. These are the presence of 'edge effects' (spurious changes), if the data are sparse near the ends, or non-uniform smoothing, if there are unequal numbers of observations throughout the age range [25].

The LMS method has been increasingly used in recent years and it was the chosen procedure for creating the 2000 CDC Growth Charts for the United States [2]. The introduction of specially designed programs (LMSChartmaker by Cole and Pan [39]) and packages for general statistical programs (package lmsqreg by Carey [40] for R) made the method rather accessible.

A proposed extension to the LMS approach, the LMSP method is even more flexible as it takes into account the presence of kurtosis in the distribution.

The main advantage of the HRY method is that no assumption about the nature of distribution is made. The approach is flexible and capable of handling many patterns of growth, with the suggestions of Pan et al. [33] and Goldstein and Pan [34] making it even more so. Nevertheless, the choice of degrees of polynomials requires considerable experience and trial, and it is not always clear how to improve the fit. Although the formulae are available, the estimation of the Z-score and centile value for further observations is not simple, unless a very basic model has been fitted. The method is vulnerable to outliers and requires specially written software, as well. Non-parametric approaches based on quantile regression are very robust, extremely flexible, provide a much better fit to the data than other methods [38] and moreover, are fully unsupervised. Although more detailed insight into the issue of quantile regressions is not so easy and straightforward compared to previous approaches, the possibilities of applications are very wide. The method has potential of identifying the features in the data undetected by other methods and is rapidly entering mainstream statistics [38]. Another compelling motivation for the quantile regression approach is the ability to extend the conventional unconditional models depending only on the age to models that incorporate prior growth and other covariates. Further, the number of implementations in statistical software is growing (e.g. package quantreg by Koenker [41] for R).

However, some drawbacks exist. The approach lacks an explicit formula to convert measurement into quantile and

Z-score. The produced curves may be irregular near the extremes, and are generally less aesthetically pleasing [38].

5. Conclusion

Cole remarked [23] that fitting smooth centile curves has always been something of a black art. We have presented several different approaches how to deal with the constructing centile charts of the fetal size.

A closer understanding how the particular method works can help us to reveal its potential and advantages and lead us to choose the most appropriate one for our goal. On the other hand, knowledge of its limitations may prevent us from misinterpretation of the results and consequently from false clinical conclusions and suboptimal care.

The explicit formula that allows one to convert a measurement into quantile or Z-score was one of the requirements set forth by a WHO expert committee [13]. This would exclude the empirical methods such as the HRY method and quantile regression approach. Hence, the choice is basically left to the trade-off between the simplicity and usability of more liminary parametric approaches, and the more flexible and applicable but less user-friendly models provided by the LMS, respectively LMSP methods. The potential of quantile regression models in the field of fetal medicine remains to be further explored.

Acknowledgements

The work was supported by the grant SVV-2010-265513.

References

- [1] Altman D.G., Chitty L.S.: Design and analysis of studies to derive charts of fetal size. *Ultrasound Obstet Gynecol* 1993; 3: 378-384.
- [2] Kuczmarski R.J., Ogden C.L., Guo S.S.: 2000 CDC growth charts for the United States: Methods and development. Hyattsville: National Center for Health Statistics; 2002.
- [3] Snijders R.J., Sherrod C., Gosden C.M., Nicolaides KH. Fetal growth retardation: associated malformations and chromosomal abnormalities. *Am J Obstet Gynecol* 1993; 168: 547-555.
- [4] Royston P., Wright E.M.: How to construct 'normal ranges' for fetal variables.

- Ultrasound Obstet Gynecol* 1998; 11: 30-38.
- [5] Sokol R.J., Chik L., Dombrowski M.P., Zador I.E.: Correctly indentifying the macrosomic fetus: improving ultrasonography-based prediction. *Am J Obstet Gynecol* 2000; 182: 1489-1495.
- [6] Wright E.M., Royston P.: A comparison of statistical methods for age-related reference intervals. *J R Statist Soc A* 1997; 160: 47-69.
- [7] Altman D.G., Chitty L.S.: Charts of fetal size: 1. Methodology. *Br J Obstet Gynecol* 1994; 101: 29-34.
- [8] Salomon L.J.: Early fetal growth: concepts and pitfalls. *Ultrasound Obstet Gynecol* 2010; 35: 385-389.
- [9] Royston P., Altman D.G.: Design and analysis of longitudinal studies of fetal size. *Ultrasound Obstet Gynecol* 1995; 6: 307-312.
- [10] Royston P.: Calculation of unconditional and conditional reference intervals for foetal size and growth from longitudinal studies. *Stat Med* 1995; 14: 1417-1436.
- [11] Royston P.: Constructing time-specific reference ranges. *Stat Med* 1991; 10: 675-690.
- [12] Waterlow J.C., Buzina R., Keller W., Lane J.M., Nichaman M.Z., Tanner J.M.: The presentation and use of height and weight data for comparing nutritional status of groups of children under the age of 10 years. *WHO Bulletin* 1977; 55: 489-498.
- [13] WHO. Physical status: the use and interpretation of anthropometry. Geneva: World Health Organization; 1995.
- [14] Salomon L.J., Bernard J.P., Duyme M., Buvat I., Ville Y.: The impact on choice of reference charts and equations on the assessment of fetal biometry. *Ultrasound Obstet Gynecol* 2005; 25: 559-565.
- [15] Salomon L.J., Bernard J.P., Ville Y.: Analysis of Z-score distribution for the quality control of fetal ultrasound measurements at 20-24 weeks. *Ultrasound Obstet Gynecol* 2005; 26: 750-754.
- [16] Altman D.G.: Construction of age-related reference centiles using absolute residuals. *Stat Med* 1993; 12: 917-924.
- [17] Royston P., Altman D.G.: Regression using fractional polynomials of continuous covariates: parsimonious parametric modelling. *Appl Statist* 1994; 43: 429-467.
- [18] Bland J.M.: The half-normal distribution method for measurement error: two case studies. Talk presented in the Statistics and Econometrics seminar series, University of York, February 2005. <http://www-users.york.ac.uk/~mb55/talks/halfnor.pdf> [Accessed 9 October 2010]
- [19] Aitkin M.A.: Modelling variance heterogeneity in normal regression using GLIM. *Appl Statist* 1987; 36: 332-339.
- [20] Shapiro S.S., Wilk M.B.: An analysis of variance test for Normality (complete samples). *Biometrika* 1965; 52: 591-611.

- [21] Shapiro S.S., Francia R.S.: An approximate analysis of variance test for normality. *J Am Statist Assoc* 1972; 67: 215-216.
- [22] McCullagh P., Nelder J.A.: Generalized linear models (2nd ed). London: Chapman&Hall/CRC; 1989.
- [23] Cole T.J.: Fitting smoothed centile curves to reference data. *J R Statist Soc A* 1988; 151: 385-418.
- [24] Cole T.J.: The LMS method for constructing normalized growth standards. *Eur J Clin nutr* 1990; 44: 45-60.
- [25] Cole T.J., Green P.J.: Smoothing reference centile curves: the LMS method and penalized likelihood. *Stat Med* 1992; 11: 1305-1319.
- [26] Box G.E.P., Cox D.R.: An analysis of transformation. *J R Statist Soc B* 1964; 26: 211-252.
- [27] Buuren S., Fredriks M.: Worm plot: a simple diagnostic device for modeling growth reference curves. *Stat Med* 2001; 20: 1259-1277.
- [28] Wilk M.B., Gnanadesikan R.: Probability plotting methods for the analysis of data. *Biometrika* 1968; 55: 1-17.
- [29] Rigby R.A., Stasinopoulos D.M.: Smooth centile curves for skew and kurtotic data modelled using the Box-Cox power exponential distribution. *Stat Med* 2004; 23: 3053-3076.
- [30] Healy M.J.R., Rabash J., Yang M.: Distribution-free estimation of age-related centiles. *Ann Hum Biol* 1988; 15: 17-22.
- [31] Cleveland W.S.: Robust locally weighted regression and smoothing scatterplots. *J Am Stat Assoc* 1979; 74: 829-836.
- [32] Silverwood R.J., Cole T.J.: Statistical methods for constructing gestational age-related reference intervals and centile charts for fetal size. *Ultrasound Obstet Gynecol* 2007; 29: 6-13.
- [33] Pan H.Q., Goldstein H., Yang Q.: Nonparametric estimation of age-related centiles over wide age ranges. *Ann Hum Biol* 1990; 17: 475-481.
- [34] Goldstein H., Pan H.: Percentile smoothing using piecewise polynomials, with covariates. *Biometrics* 1992; 48: 1057-1068.
- [35] Koenker R.: Quantile regression. Cambridge: Cambridge University Press; 2005.
- [36] Gannoun A., Girard S., Guinot C., Saracco J.: Reference curves based on non-parametric quantile regression. *Stat Med* 2002; 21: 3119-3135.
- [37] Wei Y., Pere A., Koenker R., He X.: Quantile regression methods for reference growth charts. *Stat Med* 2006; 25: 1369/1382.
- [38] Buuren S.: Worm plot to diagnose fit in quantile regression. *Stat Med* 2007; 7: 363-376.
- [39] Cole T.J., Pan H.: LMSChartmaker. <http://homepage.mac.com/tjcole> [Accessed 13 September 2010].
- [40] Carey V.: LMSqreg: An R package for Cole-Green reference centile curves. <http://www.biostat.harvard.edu/~carey> [Accessed 13 September 2010]
- [41] Koenker R.: Quantreg: An R package for quantile regression and related methods. <http://cran.r-project.org> [Accessed 13 September 2010]

Contact

MUDr. Martin Hynek

Gennet

Center for Fetal Medicine

Kostelní 9

170 00 Prague 7

Czech Republic

e-mail: martin.hynek@gmail.cz

Infrastructure for Data Storage and Computation in Biomedical Research

T. Kulhánek¹

¹Institute of Pathophysiology, 1st Faculty of Medicine of Charles University, Laboratory of Biocybernetics and Computer Aided Teaching, Prague, Czech Republic

Supervisor: Ing. Milan Šárek, Csc.

Summary

Infrastructure as a service (infrastructure which is offered to customer in the form of service of the provider) is a deployment model which allows utilize data and computing capacity of a cloud as a set of virtual devices and virtualized machines. Infrastructure as a service can be offered separately to each project. The same capacity of connected physical machines and devices can be shared. Currently, the concept of an Infrastructure as a service is tested on several projects within activity of the CESNET association, First Faculty of Medicine, Charles University, Prague and Musical and Dance Faculty of Academy of Performing Arts in Prague.

The current research in the field of computation physiology is demanding on a high computation capacity. The computation tasks are distributed to computers, which are provided by the infrastructure. The project in the field of the analysis of a human voice is demanding on high throughput of a computer network between an acoustic or video device on the local side and an analytic application on the remote high performance server side. This paper describes features and main challenges for infrastructure dedicated for such a type of an application. Infrastructure as a deployment model of cloud computing might be beneficial for a multi domain team and for collaboration and integration of a high specialized software application.

Keywords: cloud computing, infrastructure as a service, virtualization, computation physiology, identification of physiological systems, validation of physiological models, remote desktop protocol, grid computing, voice range profile

1. Introduction

Several tasks can be found in the field of an application dedicated to support biomedical research of the current distributed computing systems. The main tasks cover exchanging, storing and retrieving data. The other task is to support the analysis of data and allow long lasting parallel computation. The requirement to keep privacy of patients data is the important feature of these systems and thus it must be ensured that only authorized users may access to application and data. The high level of security is a must, or an appropriate anonymization should be implemented.

Systems focused on data exchange among different organizations try to optimize data flow via the computer network, they encrypt the data which are sent via the network, ensure the required level of reliability and integrates several incompatible systems.

The first example of distributed systems are PACS (Picture Archiving and Communication Systems). DICOM (Digital Imaging and Communications in Medicine) is the most often used standard of format and protocol to exchange medical imaging information in the field of radiology. The security of transferred images is kept on other levels of the system. PACS systems are built upon the DICOM and solves the storage and maintenance of medical images. These systems are mainly deployed and closed within a hospital or within a network of hospitals maintained by the same owner. There became a requirement to join these PACS systems from different locations. Although the DICOM protocol is used, PACS systems have proprietary implementation of

management and maintenance of DICOM images and there appeared an issue caused by incompatibilities of PACS systems.[1]. There were introduced systems for exchanging DICOM, which followed the classical structure of central storage and distributed user access (e.g. MeDiMed)[2]. There were introduced systems built as a communication centers with ability to send data among the different PACS systems (e.g. ePACS[3] or ReDiMed).

The project R-Bay was the another example of distributed systems for medicine. There were researched the possibilities how to join general systems, including exchange of DICOM images among institutions from different European countries. There was researched also the ability to provide and consume services of radiologists remotely on an international level.[4]

The systems which use computation and/or data grid is the other example of the distributed system in medicine. The project Globus Medicus is built upon the grid middleware Globus Toolkit. It provides services which presents a the DICOM interface to the user. The usage of grid middleware brings some beneficial features like reliability, security and effective transfer of data [5].

The system built within the project FONIATR is an example of the system demanding on the data transfer rate. This system supports a phoniatic examination and provides an application for the analysis of the human voice over the remote desktop protocol (RDP) in the MS Windows platform [6].

The system built within the project IDENTIFIKACE is an example of the computation system. Computational models of human physiology are developed in the MATLAB/Simulink environment and the new models also in the Modelica language [7]. Current models cover the whole complex functional parts of human physiology and reuse published relations and schema [8]. The models are validated against the data measured on patients within the project work, this is so-called identification of physiological systems. Some of the model parameters cannot be measured thus they are estimated by optimization techniques, which are demanding on performance and take a lot of a computational time. The estimation of parameters may take several hours or days on some more complex models. There are developed techniques to parallelize the computation and distribute the computation tasks to desktop computers in the laboratory or computers in the academic grid centers [9].

It's possible to build an own infrastructure for each of the previously mentioned systems. This may be, however, demanding on money, time and human resources when purchasing, installing and configuring all the needed servers and devices.

2. Methods

The infrastructure provided in an academic environment in the form of shared computational and data clusters is one of the way how to streamline the process of building up the computational and data resources. This possibility is offered by a national or international computational and data grid. Some of the projects may require a specific environment or a specific version of software library, which is not present in general grid systems. This requirement might be solved by using virtualization and cloud computing.

Computation and data grid

The computational grid is a hardware and software infrastructure that provides dependable, consistent, pervasive, and inexpensive access to high-end computational capabilities [10]. Data grid may be characterized as an integrating

architecture to do an efficient and reliable execution of data queries, which requires careful management of terabyte caches, gigabit per second data transfer over wide area networks, coscheduling of data transfer and supercomputer computation, accurate performance estimations to guide the selection of data replicas and other other advanced techniques that collectively maximize use of scarce storage, networking and computing resources [11].

It needs an additional effort to administer and maintain the grid infrastructure. This task is typically provided by a national grid initiative and the grid infrastructure is shared among different independent users. The national grid initiative was established in the Czech Republic and is maintained by the METACENTRUM activity which is one activity of the NREN (national research and educational network) provider, the CESNET association [12] and coordinates also the work with NGI from neighboring countries in the European Grid Initiative (EGI).

The system needs to be customized and use the API provided by the grid middleware Globus Toolkit, gLite and others.

Desktop grid

Anybody (from the academic and research community) can join and use the grid built by the NGI. Usually only the provider of NGI maintains and enlarges this infrastructure and grid middleware.

There exists different concepts how to build the grid. Everybody can enlarge the grid infrastructure by joining their computer and decides which application can use their computer for computation. The most well known example of such a grid system is SETI@home [13]. The concept of SETI@home follows the idea that anyone connected to Internet can join a grid by downloading a small client program and execute it in the background which periodically asks for computational jobs and computes in the background or as a screen-saver. The grid nodes are typically PCs owned by individuals. Such systems are usually referred as Volunteer grid

systems or Desktop Grids and a general desktop grid system BOINC is used to build such systems, customized server site and client application to form a custom desktop grid application [14], [15].

Cloud computing

Virtualization is a technology which provides separation between a software layer and an underlying hardware layer. It allows execution of one of more so-called virtual machines sharing one physical hardware. The virtual machine is fully or partly (paravirtualization) separated from the physical layer of the hardware and thus different platforms (Windows, Linux) may work together on one physical machine concurrently. Virtualization techniques introduce some overhead when translating an isolated application instruction to the lower level of a system. However, the open source paravirtualization system XEN does not impose an onerous performance penalty comparing to non-virtualized operating system configuration [17].

The virtualization is sometimes characterized as a key technology which enables cloud computing and execution of different isolated systems on shared hardware.

Virtual infrastructure

Virtual organization is a group of users, who share the same resources. The virtual infrastructure belonging to a virtual organization is built from virtual machines connected via the network, which may be virtual too and accessible only to users from the virtual organization. Figure 1 shows an example of several virtual organizations and their infrastructures. On the right part there is a schematic view on physical connections among different organizations (hospitals, research institutions) via the academic network or the Internet. The physical resources are shown as vertexes and network connections are shown as edges. Each cloud shows one virtual infrastructure. On the left part there is a physical server executing more virtual machines, each machine belongs to a different virtual infrastructure.

3. Results

The pilot infrastructure dedicated for the medical application was established within the CESNET's activity "Application support" in several locations in Prague. It utilizes the open-source virtualization system XEN and tools of the operating system Linux for configuring virtual machines and virtual infrastructure as a service.

The instance of the grid system Globus MEDICUS was the primary system deployed to a set of virtual machines. It was shown that the grid system based on open standards can be easily integrated with current medical systems using the DICOM format [18].

The application to support a phoniatic examination was deployed next to the previous system to exchange medical images. It was needed to develop an enhancement of the RDP protocol, which adds the transfer of an audio signal from local computer's microphone to the remote application [6]. This system is currently deployed on one virtual machine. See Figure 2.

The pilot infrastructure contains also the system for identification of physiological systems, which offers a web service distributing the computational task to desktop computers connected via the desktop grid system BOINC and SZTAKI Desktop Grid API [9]. The schema in Figure 3 shows the architecture of the system. The server is in operation as an independent virtual machine and contains the web service. Some of the BOINC workers are in operation as independent virtual machines deployed on less used physical servers of the pilot infrastructure. Some of the desktop computers of laboratory and classroom of the First Faculty of Medicine are connected to this desktop grid system. Other computers may be easily joined. Current research is focused on the possibility to enhance the computational capacity of the infrastructure by the resources provided by the NGI MetaCentrum. There is also researched an utilization of GPU computing.

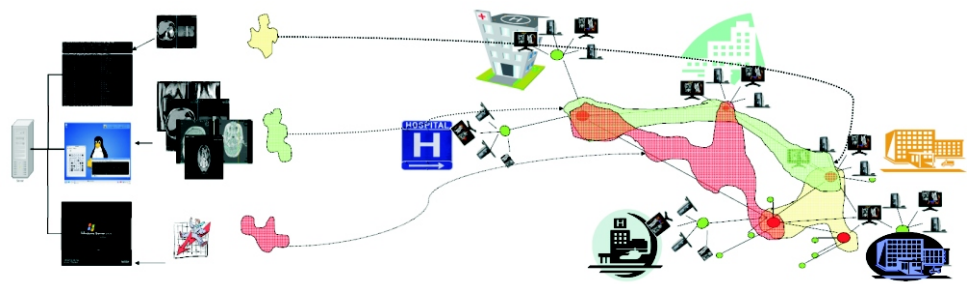


Fig. 1. General schema of virtual infrastructure.

All the physical servers of the pilot infrastructure have a virtual environment built upon the XEN system and they share the IP network addresses. Each virtual machine has its own disk partition within the Logical Volume Management (LVM) on the physical server. The virtual machines are administered by the tool virt-manager and the network environment is configured with the tool iptables.

4. Discussion

The pilot infrastructure can be characterized as a private cloud, which is accessible only to the limited community of users from the field of biomedical research. Virtual machines share the physical network connection via IPv4. Because of the lack of numbers of unique IPv4 addresses, the configuration of network

services (webserver, RDP) is realized using network address translation and port mapping. If the network devices passed to the version 6 of IP protocol, there would be opened again the possibility to provide unique IP addresses to virtual machines and there would not be needed extra configurations of network address translation and port mapping.

There are not used special tools to administer cloud within the pilot infrastructure, because the number of projects is relatively small currently. Anyway, there exist free or commercial products (Eucalyptus, OpenNebula, VMWare vSphere), which provide a set of tools to automatize the maintenance of private cloud, including virtual network configuration, live migration of virtual

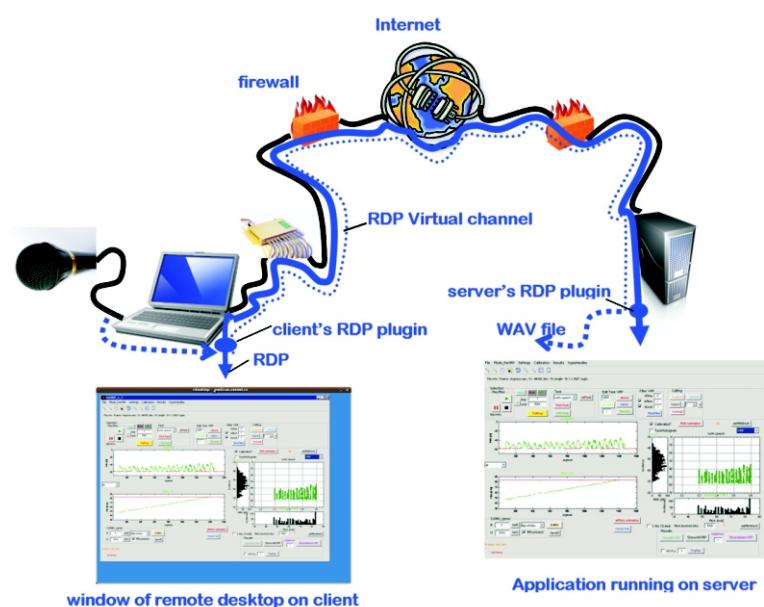


Fig. 2. Schema of system for human voice analysis and remote recording via RDP protocol.

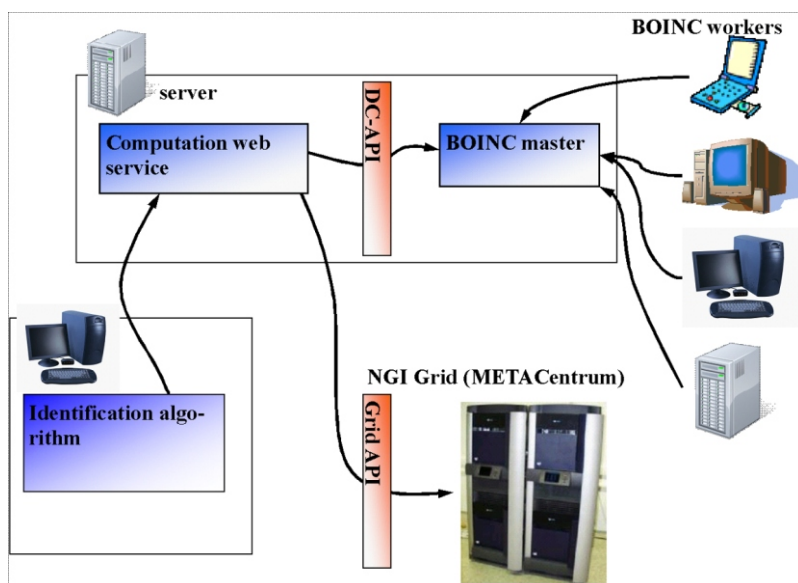


Fig. 3. Schema of computational infrastructure for identification of physiological systems.

machine, etc. Deploying these tools will be necessary in future after expected enhancement of the physical capacity, which is planned to be built within the academic environment of the Czech Republic.

The important question is: which type of the application is suitable for clouds operating on physical resources spread in different geographical locations compared to clouds operating in supercomputing centers. Clouds in supercomputing centers are suitable for highly parallel tasks which need fast communication between parallel computational tasks. Cloud operating on physical servers in different geographical locations can offer a free capacity in the time period, when the owner does not utilize its physical resources and offers them to other users of cloud.

5. Conclusion

It is possible to operate a private cloud on the physical infrastructure and to provide the virtual infrastructure to the users, who can utilize it to execute their own applications and systems. The infrastructure as a service can open an access to distributed systems to a higher amount of users, who have been so far prevented from using them by a complicated

administration, too long process of purchasing and installing computing resources.

The cloud operating on physical servers in different geographical locations can be a suitable complement to the clouds in supercomputing centers.

Acknowledgment

The work was supported by the grant SVV-2010-265 513.

References

- [1] Kulhánek T., Šárek M.: Virtualizace a integrace v gridovém PACS systému, MEFANET 2008, Brno, ISBN 978-80-7392-065-4
- [2] Dostal, O., Javornik, M., and Ventruba, P.: "Collaborative environment supporting research and education in the area of medical image information". International Journal of Computer Assisted Radiology and Surgery. Springer Science+ Business Media. 1. 98. 2006.
- [3] E-PACS, DICOM komunikace mezi zdravotnickými zařízeními, 2010, czech language, web: <http://www.epacs.cz/>
- [4] Pedersen C. D., Rasmussen J., Beck C.: R-Bay; Creating an eMarketplace for the transfer of imaging related eHealth services in Europe. Final project report, version 2, web: <http://www.r-bay.org>
- [5] Erberich S.G., Silverstein J.C., Chervenak, A., Schuler, R., Nelson, M.D., and Kesselman, C.: "Globus MEDICUS-Federation of DICOM Medical Imaging Devices into Healthcare Grids". Studies in Health Technology and Informatics. IOS Press; 1999. 126. 269. 2007.
- [6] Kulhánek T., Frič M., Šárek M.: Remote Analysis of Human Voice Lossless Sound Recording Redirection, Proceedings of Biosignal 2010: Analysis of Biomedical Signals and Images, ISSN 1211-412X, ISBN 978-80-214-4106-4
- [7] Stodulka, P., Privitzer, P., Kofránek, J., Tribula, M., and Vacek, O.: Development of Web Accessible Medical Educational Simulators". Proceedings of EUROSIM 2007, Ljubljana. 2. 16.
- [8] Abram S.R., Hodnett B. L., Summers R.L., Coleman T.G., Hester R.L.: Quantitative circulatory physiology: an integrative mathematical model of human physiology for medical education. Adv. Physiol. Educ., 31, 202-210, (2007).
- [9] Kulhánek T., Šilar J., Mateják M., Privitzer P., Kofránek J., Tribula M.: Distributed computation and parameter estimation in identification of physiological systems, VPH conference 2010, Brussels
- [10] Foster I., Kesselman C. "Computational grids", Vector and Parallel Processing, VECPAR 2000, 2001 Springer
- [11] Chervenak A., Foster I., Kesselman C., Salisbury C., Tuecke S.: The Data Grid: Towards an Architecture for the Distributed Management and Analysis of Large Scientific Datasets. Journal of Network and Computer Applications, Elsevier, 2000
- [12] Šustr Z., Sitera J., Mulač M., Ruda M., Antoš D., Hejtmánek L., Holub P., Salvat Z., Matyska L.: MetaCentrum, the Czech Virtualized NGI, 2009, web: <http://egee.cesnet.cz>
- [13] Anderson D. P., Cobb J., Korpela E., Lebofsky M., and Werthimer, D. 2002. SETI@home: an experiment in public-resource computing. Commun. ACM 45, 11 (Nov. 2002), 56-61.
- [14] Anderson D. P. 2004. BOINC: A System for Public-Resource Computing and Storage. In Proceedings of the 5th IEEE/ACM international Workshop on Grid Computing (November 08 - 08, 2004). International Conference on Grid Computing. IEEE Computer Society, Washington, DC, 4-10.
- [15] Balaton Z., Gombás G., Kacsuk P., Kornafeld A., Kovács J., Marosi C.A., Vida G., Podhorszki N., Kiss T.: SZTAKI Desktop Grid: a Modular and Scalable Way of Building Large Computing Grids. Proceedings of the 21st International Parallel and Distributed Processing Symposium (IPDPS 2007), 26-30 March 2007, Long Beach, California, USA

- [16] Velte T., Velte A., and Elsenpeter R.: Cloud Computing: A Practical Approach". McGraw-Hill Osborne Media. 2009.
- [17] Youseff L., Wolski R., Gorda B., Krintz, C. Paravirtualization for HPC Systems. Frontiers of High Performance Computing and Networking " ISPA 2006 Workshops. 474486. 2006.
- [18] Kulhánek T., Šárek M.: Processing of medical images in virtual distributed environment. ACM. 13. 2009.

Contact

Mgr. Tomáš Kulhánek

Institute of Pathophysiology,
First Faculty of Medicine of Charles
University
Laboratory of Biocybernetics and
Computer Aided Teaching
U Nemocnice 5
128 53 Prague 2
Czech Republic
e-mail: tomaton@centrum.cz

Language of Czech Medical Reports and Classification Systems in Medicine

P. Přečková¹

¹Centre of Biomedical Informatics, Institute of Computer Science AS CR, Prague, Czech Republic

Supervisor: Prof. RNDr. Jana Zvárová, DrSc.

Summary

The objective of the paper is to compare Czech medical reports written in a free text and by means of a software application; to analyze the usability of international classification systems in the Czech healthcare environment. The analysis of medical reports was based on the attributes of the Minimal Data Model for Cardiology (MDMC). We have used medical reports written in a free text and medical reports from the ADAMEK software application where data are stored in a structured way. For our work SNOMED CT and ICD-10 have been used. We have focused on the language of Czech medical reports and the application of aforementioned international classification systems in MDMC. We have compared how well attributes of MDMC are recorded in textual medical reports and in medical reports recorded structurally by means of the ADAMEK software application. We have made the language analysis of the Czech textual medical reports. We compared how MDMC attributes are recorded in the ADAMEK application and in medical reports written in a free text. To conclude, using a free text in medical reports is very inhomogeneous and not standardized. The standardized terminology would bring benefits to physicians, patients, administrators, software developers and payers. It would help healthcare providers that it could provide complete and easily accessible information that belongs to the process of healthcare and it would result in better care of patients. The use of international classification systems is a necessary first step to enable interoperability of heterogeneous electronic health records.

Keywords: terminology, synonyms, classification systems, thesaurus, nomenclature, electronic health record,

interoperability, semantic interoperability, cardiology, atherosclerosis

1. Introduction

Determination, denomination, and classification of medical terms are not optimal. The proof is that for one term we can often meet with more than ten synonyms. This problem has intensified with an introduction of a computer technology to healthcare. Using computers means higher uniqueness of data feeding, of term definitions, their precise denomination, etc., thereby the significant drawback becomes more noticeable.

Generally, in the scientific terminology it is more advantageous to use only one expression for one term. Computers are able to learn synonyms but it enlarges dictionary databases and the number of necessary operations grows. Moreover, synonymy in the scientific terminology leads to inaccuracy and misunderstanding. In current medicine we can meet with many synonyms for one single disease. That was the reason why coding systems providing codes for any medical findings have arisen.

Nowadays, there is a big boom in the development of electronic health records (EHR). There is a general agreement that electronic, computer-based medical records have the potential to improve the quality of medical care [1]. Within the concept of the EHR, the patient is understood as an active partner who is accessing, adding and managing health-related data [2].

Safe and appropriate clinical information exchange among various electronic health records is essential for continuity of patients' care in various times, at various places and in various healthcare providers. Mapping of electronic health record

attributes to various classification systems is a very important step for semantic interoperability among these various systems. EHRs and semantic interoperability are very topical issues and they are discussed in many papers [3], [4], [5], [6]. Semantic interoperability based on the Czech language has been studied in [7], [8], [9], [10], [11].

Clinical data from electronic health records has traditionally contained a small proportion of fixed field data (often obtained from pick lists) and larger quantities of a free text [12]. In our paper we will discuss both of these types of data.

2. Coding and classification systems

Coding systems limit the variability of expression. Only the approved terms and their phrases can be used according to strictly given rules. Formal codes are usually used instead of the approved terms. In many cases it is useful if coding systems show also not approved terms, which are often used as synonyms for approved terms.

Let us show you some of the most widespread international classification systems.

2.1 ICD International Classification of Diseases and Related Health Problems

The foundation of the International Classification of Diseases was laid by William Farr in the year 1855. The World Health Organization took it over in the year 1948. At that time it was its 6th revision. Since 1994 the 10th revision of ICD is in use and it contains 22 chapters. ICD has become an international standard for a classification of diseases and for many epidemiological and management needs in healthcare.

These include a general situation of health in different population groups and monitoring of the incidence and prevalence of various diseases and other health problems in relation to other variables. It is used to classify diseases and other health problems that are recorded in many types of health records, including death certificates and hospital records [13]. ICD is available in six official languages of WHO and in other 36 languages, including Czech.

2.2 SNOMED CT

SNOMED Clinical Terms originated from two terminologies: SNOMED RT and Clinical Terms Version 3 (Read Codes CTV3). SNOMED CT represents the Systematized Nomenclature of Medicine Reference Terminology developed by the College of American Pathologists. It serves as a common reference terminology for gathering and acquiring health data recorded by organizations or individuals. The Clinical Terms Version 3 was developed by the United Kingdom's National Health Service in the year 1980 as a mechanism for storing structured information on primary care in Great Britain. These two terminologies united in the year 1999 and a highly complex terminology SNOMED CT arose [14], [15], [16]. Nowadays we can meet with American, British, Spanish, and German versions of SNOMED CT.

2.3 MeSH

Medical Subject Headings [17] is a vocabulary controlled by the National Library of Medicine (NLM). It is composed of terms, which denominate keywords hierarchically and this hierarchy helps with searching on various levels of specificity. Keywords are arranged not only alphabetically but also hierarchically. NLM uses MeSH for indexing of papers from world best biomedical journals for the MEDLINE/PubMED database. MeSH is used also for a database cataloguing books, documents, and audiovisual materials. Each bibliographical reference is connected with a class of terms in the MeSH classification system. Searching inquiries use also the MeSH vocabulary to find papers with required topics. The MeSH vocabulary is updated continuously and it is also controlled by specialists creating it. They collect new terms appearing in

scientific literature or in the arising fields of research. They define these terms in the frame of the contents of the existing vocabulary and they recommend their adding to the MeSH vocabulary. There exists also the Czech translation of MeSH.

2.4 LOINC

Logical Observations Identifiers Names and Codes (LOINC) [18] is a clinical terminology, which is important for laboratory tests and laboratory results. In the year 1999 the HL7 organization accepted LOINC as a preferred coding system for names of laboratory tests and clinical observations

2.5 ICD-O

The ICD-O classification system [19] is an extension of the International classification of diseases for oncology coding. It is a four-dimensional system. The dimensions are appointed to classify morphological kinds of tumors. The third version of ICD-O is used nowadays.

2.6 TNM

The TNM classification system [20] is a clinical classification of malignant tumors used for comparison of therapeutic studies. The TNM system is based on the assessment of three components: T - the extent of the primary tumor; N - the absence or presence and extent of regional lymph node metastasis and M - the absence or presence of distant metastasis. It proceeds from the knowledge that, for the disease prognosis, the localization and spread of a tumor is the most important.

2.7 Other Classification Systems

Currently, there are more than one hundred of various classification systems. These are for example *AI/RHEUM*; *Alternative Billing Concepts (ABC)*; *Alcohol and Other Drug Thesaurus (AOD)*; *Beth Israel Vocabulary*; *Canonical Clinical Problem Statement System (CCPSS)*; *Current Dental Terminology (CDT)*; *DSM (Diagnostic and Statistical Manual of Mental Disorder)*; *Medical Entities Dictionary (MED)*; *Current Procedural Terminology (CPT)*; *International Classification of Primary Care (ICPC)*; *McMaster University Epidemiology Terms*; *CRISP Thesaurus*; *Coding Symbols for a Thesaurus of Adverse Reaction Terms*

(*COSTART*); *Diseases Database*; *DXplain*; *Gene Ontology (GO)*; *Healthcare Common Procedure Coding System (HCPCS)*; *Home Health Care Classification (HHCC)*; *Health Level Seven Vocabulary (HL7)*; *Master Drug Data Base (MDDDB)*; *Medical Dictionary for Regulatory Activities Terminology (MedDRA)*; *Multum MediSource Lexicon (MMSL)*; *NANDA nursing diagnoses*; *NCBI Taxonomy* and many others.

3.1 Conversion tools

The increasing number of classification systems and nomenclatures requires designing of various conversion tools for transfer among main classification systems and for recording of relations among terms in these systems. Extensive ontologies and semantic networks are modeled for information transfer among various databases. Metathesauri are designed to monitor and connect information from various heterogeneous sources.

The Unified Medical Language System (UMLS) [21] was initiated in the year 1986 in the National Library of Medicine in the USA. UMLS knowledge sources are universal. It means they are not optimized for individual applications. It is an intelligent automated system, which "understands" biomedical terms and their relations and it uses this understanding for reading and organization of information from machine processed sources. Its aim is to compensate terminological and coding differences of these non-homogeneous systems and also language varieties of users. It is a multilingual thesaurus of classification systems on a high-capacity medium, which enables to transfer coded terms among various classification systems.

UMLS is based on three knowledge sources: Metathesaurus, Semantic Network, and SPECIALIST Lexicon. The Semantic Network contains information about semantic types and their relations. The SPECIALIST Lexicon records syntactic, morphologic, and orthographic information of each word or a term.

The UMLS Metathesaurus is an extensive, multi-purpose, and multilingual database. It contains information about biomedical, healthcare and their relative terms, their various expressions and relations among them. Its main aim is to connect alternative expressions of the same terms and to identify useful relations among various terms. If thesauri use the same expressions for different terms, then both meanings are present in the Metathesaurus and we can also see which meaning is used in which thesaurus. If the same term is used in different hierarchical contexts in various thesauri, then the Metathesaurus keeps all these hierarchies. The Metathesaurus does not give one consistent view, but it keeps many views, which are present in source thesauri.

4. Minimal data model for cardiology

Since the year 2000 our department has been involved in the research of the joint workplace consisting of two universities, two hospitals and the Institute of Computer Science AS CR. This joint workplace is called the EuroMISE Centre and it runs also two outpatient departments of preventive cardiology. In the year 2002 the Minimal Data Model for Cardiology (MDMC) was developed [22], [23]. MDMC is a set of approximately 150 attributes, their mutual relations, integrity restrictions, units, etc. Prominent professionals in the field of Czech cardiology agreed on these attributes as on the basic data necessary for an examination of a patient in cardiology.

As cardiology is a very extensive field MDMC is limited to atherosclerotic cardiovascular diseases. The aim of MDMC has been to form a minimal set of attributes, which are needed to be observed in all patients from the cardiological point of view so a patient could be ranked into ill persons or persons in risk from the perspective of cardiovascular diseases.

Necessary administrative attributes are one part of MDMC. Nevertheless, they are not included in our analysis as the

analyzed medical reports have been anonymous and therefore without any administrative data.

MDMC consists of eight groups of attributes. After the *administrative part* there is a family history part consisting of information about mother, father and siblings. The next part is the *social history and addiction* focusing on the marital status, physical activities, mental stress, levels of smoking and alcohol consumption rates. One part of MDMC is devoted to allergies, mainly to drug *allergies*. The *personal history part* detects the presence of diabetes mellitus, there is observed whether the patient suffered from a stroke, whether he/she is treated with an ischemic disease of periphery arteries, there are attributes related to aortic aneurysm, other relevant diseases and menopause in women. In the part of MDMC called *Current difficulties of a possible cardiological origin* physicians are focused on the shortness of breath, chest pain, palpitations, swellings, syncope, cough, hemoptysis, and claudication. Another part of MDMC determines what kind of a *treatment* the patient undergoes, what type of a diet is prescribed and which medications he/she uses. In the part of the *physical examination*, the patient's weight, height, body temperature, BMI, WHR, blood pressure, pulse and breathing rates and pathological finding are determined. *Laboratory testing* is focused on blood glucose, uric acid, total cholesterol, HDL-cholesterol, LDL-cholesterol and triacylglycerols. The last part of the MDMC is focused on attributes related to ECG. The beat frequency, the average PQ and QRS intervals and the full description of ECG is described there.

MDMC has become a basis for the ADAMEK software application where data are stored in a structured way. After its completion the data collection started in March 2002. The data were and still are collected in two outpatient departments of preventive cardiology of the EuroMISE Centre. To April 1st, 2010 there were data about 1189 patients.

5. The language of Czech medical reports and the application of international classification systems in MDMC

5.1 The Czech language

The Czech language belongs to the western group of Slavic languages. As a Slavic language Czech belongs to the eastern, or satem, division of Indo-European languages. The Czech language separated itself from other Slavic languages by a number of changes, most of which took place in the 10th through 16th centuries. By the end of the 15th century the Czech language had virtually lost the dual number and two of the Slavic past tenses. On the other hand, the role of the verbal aspect had grown more significant and the number of declensions had increased. At the beginning of the 15th century the religious reformer Jan Hus (John Huss) devised a diacritical writing system, placing diacritical marks over some Latin letters to distinguish the Czech palatal/palatalized consonants (č, ď, ň, ř, š, ť, ž) and long vowels (á, é, í, ó, ú, ý). In the 16th century the letter "ů", indicating the long "u", was added. The only digraph surviving in modern Czech is "ch".

The Czech language has about 10 million speakers in the Czech Republic and about 200 000 speakers in other countries. These are mostly emigrants and children of emigrants who left the country in sizable migrations around World War I, World War II, and the year 1948 and 1968. Many Czech speakers can be found especially in Austria (mostly in Vienna), Poland, Germany, Ukraine, Croatia, in Western Romania, in Australia, and Canada. However, the largest group of Czech speakers outside the Czech Republic lives in the United States of America, in cities like New York, Chicago and in a number of rural communities in Texas, Wisconsin, Minnesota, and Nebraska.

The Czech language belongs to the free constituent order languages [24].

5.2 The language analysis of the Czech medical reports written in a free text

The style of writing medical reports is not standardized. We can find differences not

only in medical reports from various physicians but also individual medical doctors write the same concepts in different forms. In the following part we will concentrate on linguistic and lexical differences in Czech medical reports written in a free text.

Diacritic: As we have formerly stated the Czech language uses the diacritical writing system. As an example of diacritical letters let us mention for example letters “ě, č, ř, ž,” and others. However, it is faster for physicians to write without these diacritical marks and use letters “e, c, r, z”. Such a text is for Czech native speakers understandable but it is difficult for computational processing.

Typing errors: Typing errors represent a bigger problem and they are very frequent. The text is then very hardly usable for computational processing.

Spaces: A similar problem is spaces omitting between words, which results in merging of two words in one and this is again unusable for computational processing. Physicians vary also in using spaces in front of units. We may find a form with a space, e.g. “2.5 mg” but also the form without a space, e.g. “4mg”. The same situation is with attributes using a slash. Some physicians use the variant without a space, e.g. “80/min”, the others use a space, e.g. “70 / min”.

Figure 0: For computational processing it is difficult if some physicians use instead of the figure 0 the capital letter O.

Abbreviations: As physicians are often pressed of time they abbreviate words while writing medical records. The problem here is that there exists not a single rule how particular attributes should be abbreviated. Therefore we can often meet diversely abbreviated same words. We can find also examples that in one medical record written by one physician one attribute is abbreviated several times and each time differently. With abbreviated words the problem of full stop omitting behind the abbreviated word arises.

Rounding-off: Another part where we can find many discrepancies is connected with numerical values. We can meet with that

one physician rounds one attribute to integers, while another physician rounds the same attribute with the precision of one or two decimal numbers. Sometimes numerical values are presented as ranges, e.g. “70-80”. Often only an approximate indication is entered, e.g. “diastolic pressure around 70”. Some attributes are not expressed in numbers but in words, e.g. “blood pressure is within the normal range”.

Arabic and Roman numerals: There is a difference in usage of Arabic and Roman numerals. For example heart sounds may be found in both ways “heart sounds 2” and “heart sounds II”.

Synonyms: The Czech language is very rich in synonyms and they are highly used also in medical reports.

Orthography: Some physicians use a newer version of spelling, the others the older one.

Time data: Recording of time is not standardized as well. In medical reports we can meet the name of the month, e.g. “February 2006” but also the month order, e.g. “2/2006”.

Drugs administering: In the medical reports there appear a lot of various ways how to describe the time when a patient should administer a drug (e.g. 1-0-0 vs. 1 pill in the morning vs. 1 in the morning vs. 1x in the morning).

Attribute values: The same values of the attributes are often recorded in different ways. For example: diabetes mellitus can be found in a form of diabet, diabet., diabetes mellitus of the 2nd type, diabetes mellitus of II type on a diet, DM of the 2nd type.

These are not problems only of writing medical reports but the same orthographic errors may be found e.g. in web pages [25].

5.3 The analysis of MDMC attributes in medical reports written in a free text
The analysis of medical reports written in a free text was based on the attributes of MDMC. Medical reports were anonymous and therefore administrative data were not possible to analyze.

If we look at individual attributes, we can see that only *diastolic* and *systolic pressures* were recorded in all textual medical reports that have been analyzed. In 96.30 % of the textual medical reports *medications* that a patient uses or a physician administers are recorded. Also the *weight* is recorded in 96.30 % of medical reports. In contrast, the *height* is recorded only in 74.07 % of reports. In 27 analyzed reports some of the MDMC attributes have not appeared not even in one case. This includes the following attributes: *aortic aneurysm*, *angina pectoris*, *ischemic disease of lower limbs*, *when DM was discovered*, *silent ischemia*, *body temperature* and others. *Drug allergy* is mentioned in 22.22 % of reports, whether a patient drinks or does not drink an *alcohol* in 51.85 % reports, *chest pain* in 37.04 %. The *total psychical stress* is recorded in 11.11 % of reports, *physical load in a job* in 11.11 % of reports, *total cholesterol* in 70.37 %, *drinking of a black coffee* in 22.22 %, *various other examinations* in 62.96 %. The *breathing frequency* has been found in 3.7 % of reports, *diabetes mellitus* in 40.74 %, *diet* in 59.26 %, *glycaemia* in 51.85 %, *HDL cholesterol* in 66.67 %. The presence or absence of *hypertension* has been recorded in 70.37 %, *left ventricular hypertrophy* in 11.11 %, *myocardial infarction* in 14.81 %, *the average amount of cigarettes in a smoker* in 51.85 % of reports and so on.

We can say that while recording results of examinations by means of a free text a lot of attributes is left unrecorded. There may be several reasons for that. Physicians do not have a strictly given skeleton according which they should proceed and it can happen that they may forget some attributes. In software applications it is not possible because if physicians forget to fill in the given attribute, the application will not let them to continue. Another reason why some attributes are not recorded may be the fact that physicians from the previous attributes know that the next attribute cannot be present and that is the reason why they do not ask about it and they do not record it.

But from the free text medical report we do not know whether these missing attributes have been checked or whether they have been deduced by physicians on the basis of previous knowledge.

5.4 The analysis of MDMC attributes in the ADAMEK software application

For this analysis 1118 medical reports from the outpatient department of preventive cardiology of the EuroMISE Centre have been used.

In all medical reports generated by ADAMEK, in 100 %, these MDMC attributes were recorded: *drug allergy*, *aneurysm of aorta*, *angina pectoris*, *chest pain*, *swelling of lower limbs*, *asthma*, *left ventricular hypertrophy*, *myocardial infarction*, *other allergies*, *cough after ACE inhibitors*, *claudication*, *silent myocardial ischemia*, *systolic pressure*, *type of DM treatment*. Even from this enumeration it is clear that by means of the software application a higher number of attributes is recorded.

If we compare how individual attributes are recorded in free text medical reports of the hospital and in ADAMEK medical reports, we reach the following results, see Table 1. *Drug allergy* is recorded in free text medical reports in 22.2 %, in the ADAMEK application in all, i.e. 100 % of reports. The answer whether a patient suffers or not from *aneurysm of aorta* has been recorded in all reports from the ADAMEK application but in none of textual medical reports. Questions about a *chest pain* have been recorded in all medical reports using the ADAMEK application but only 37.0 % of free text medical reports included remarks on a chest pain. The *level of stress* has been recorded in 96.2 % of the ADAMEK medical reports, in free text reports it was only in 11.1 %. The same percentage was reached in the attribute of *physical load in job*, in the ADAMEK application it was in 94.8 %. *Total cholesterol* has been recorded in 83.4 % of the ADAMEK medical reports, in free text reports it was 70.4 %. Presence or absence of *diabetes mellitus* was recorded in the software application in 95.9 %, in the free text reports it was 40.7 %. *Glycaemia* was recorded in 77.6 % of reports from the ADAMEK application and in 51.9 % in free text medical reports. *Cholesterol* was

recorded in 917 reports of the software application, which is 82.0 %, while in free text medical reports we can meet with this attribute recorded only in 66.7 % of analyzed reports. Both kinds of reports are the closest in *weight* that is recorded in the ADAMEK application in 97.9 % of medical reports and in 96.3 % of free text medical reports. Presence or absence of *hypertension* can be met in 95.3 % of the ADAMEK reports, while only in 70.4 % of free text reports. A big difference can be found in *left ventricular hypertrophy*. It is recorded in all medical reports from the ADAMEK application but only in 11.1 % of free text medical reports. Another big difference is e.g. in the *menopause* attribute that is recorded in 96.9 % of reports from the ADAMEK application but only in 7.4 % of free text medical reports. The PQ interval is recorded by means of the software application in 89.1 % and in 62.9 % in free text medical reports. Similarly the *QRS interval* is in ADAMEK medical reports recorded in 89.5 % and in 66.7 % in free text medical reports.

In Table 1 there is lucidly displayed the percentage of recorded selected attributes in MDMC in medical reports using the ADAMEK application and in medical reports written in a free text.

5.5 Attributes of MDMC coded by means of SNOMED CT

Table 2 shows several examples of MDMC attributes to which the ConceptID from the SNOMED classification system has been allocated. The first prerequisite for this coding is the translation of the attributes to the English language as there is not a Czech version, yet.

5.6 Attributes of MDMC coded by means of ICD-10

As ICD-10 is one of a few international medical classification systems translated to the Czech language, we have tried to code the attributes of MDMC by means of this classification. In Czech the abbreviation for this classification is MKN-10. Table 3 shows a comparison of MDMC attributes coded by means of ICD-10 (MKN-10) and SNOMED CT.

Tab. 1. Percentage of recorded values of the selected MDMC attributes in 1118 ADAMEK medical reports and in 27 free text medical reports.

MDMC attribute	ADAMEK medical reports n=1118	free text medical reports n=27	free text medical reports – 95% confidence interval	
			lower limit	upper limit
drug allergy	100.0 %	22.2 %	8.6 %	42.3 %
aneurysm of aorta	100.0 %	0.0 %	0.0 %	12.8 %
angina pectoris	100.0 %	0.0 %	0.0 %	12.8 %
chest pain	100.0 %	37.0 %	19.4 %	57.6 %
level of stress	96.2 %	11.1 %	2.4 %	29.2 %
total cholesterol	83.4 %	70.4 %	49.8 %	86.2 %
diabetes mellitus	95.9 %	40.7 %	22.4 %	61.2 %
asthma	100.0 %	55.6 %	35.3 %	74.5 %
physical load in job	94.8 %	11.1 %	2.4 %	29.2 %
glycaemia	77.6 %	51.9 %	32.0 %	71.3 %
HDL cholesterol	82.0 %	66.7 %	46.0 %	83.5 %
weight	97.9 %	96.3 %	81.0 %	99.9 %
hypertension	95.3 %	70.4 %	49.8 %	86.2 %
myocardial infarction	100.0 %	14.8 %	4.2 %	33.7 %
PQ interval	89.1 %	62.9 %	42.4 %	80.6 %
other allergies	100.0 %	18.5 %	6.3 %	38.1 %
claudication	100.0 %	3.7 %	0.9 %	19.0 %
smoker	96.5 %	66.7 %	46.0 %	83.5 %

Tab. 2. Selected attributes of MDMC coded by means of SNOMED CT.

Attributes of MDMC (in Czech)	English equivalent	SNOMED CT (Concept ID)
alergie na léky	Drug allergy (disorder)	416098002
	Allergic reaction to drug (disorder)	416093006
hypertenze	Essential hypertension (disorder)	59621000
	High blood pressure (&essential hypertension)	194757006
	Essential hypertension NOS (disorder)	266228004
ischemická choroba srdeční	Ischemic heart disease (disorder)	414545008
dušnost	Asthma (disorder)	187687003
bolest na hrudi	Dull chest pain (finding)	3368006
palpitace	(Palpitations) or (awareness of heartbeat) or (fluttering of heart)	161965005
otoky	Swelling or edema (finding)	248477007
synkopa	Syncope (disorder)	271594007
klaudikace	Claudication (finding)	275520000
hmotnost	On examination – weight NOS (finding)	162770007
	Height and weight (observable entity)	162879003
výška	Body height measure (observable entity)	50373000
tělesná teplota	Body temperature finding	105723007
	Body temperature (observable entity)	276535009
obvod pasu	Abdominal girth measurement	48094003

Tab. 3. Selected attributes of MDMC coded by means of ICD-10 (MKN-10) and SNOMED CT. (Part I. - Allergy).

Attributes of MDMC (in Czech)	Term in MKN-10	MKN-10 code	English equivalent	SNOMED CT (Concept ID)
alergie přítomna	alergie	T78.4	allergy manifested	not found
alergie na léky	alergie na lék	T88.7	drug allergy (disorder)	416098002
			allergic reaction to drug (disorder)	416093006

As the very title of the International Classification of Diseases shows this classification can be used to encode particular diseases, syndromes, pathological conditions, injuries, difficulties and other reasons for the contact with healthcare services, i.e. the type of information that is being registered by a physician. Unfortunately, using this classification we cannot encode many attributes of the Minimal Data Model for Cardiology, such as marital status, education, mental stress, physical stress, physical activity, smoking, alcohol drinking, physical examination (weight, height, body temperature, BMI, WHR, etc.), laboratory tests (total cholesterol, HDL-cholesterol) or a description of ECG. ICD-10 can be used only for the parts of MDMC related to personal history and current difficulties of a possible cardiological origin (see Table 3).

5.7 Standardization of Clinical Contents

Attributes, from the point of view of possibilities of their mapping to standard coding systems, can be classified in the following way:

- *Trouble-free attributes* - i.e. attributes, which can be mapped in a direct way, only one possibility of mapping exists, possibly there are only synonyms with exactly same meanings and therefore the same classification code (e.g. patient first name, current smoker, motility, height of a patient, etc.).
- *Partially problematic attributes* - i.e. attributes, which can be mapped in a way that there are several possibilities of mapping to different synonyms, which differ slightly in their meanings and usually in their classification codes (e.g. ischemic cerebro-vascular stroke, angina pectoris, hypertension, congestive cardiac failure, etc.).
- *Attributes with a too small granularity*, i.e. attributes describing certain characteristics on a too general level so that classification systems contain only terms of a narrower meaning (e.g. e-mail in MDMC versus e-mail to work / e-mail to home / e-mail of a physician and so on in classification systems).

Tab. 3. Selected attributes of MDMC coded by means of ICD-10 (MKN-10) and SNOMED CT. (Part II - Personal history).

Attributes of MDMC (in Czech)	Term in MKN-10	MKN-10 code	English equivalent	SNOMED CT (Concept ID)
diabetes mellitus	diabetes typu I	E10.-	diabetes mellitus type 1 (disorder)	46635009
	inzulin dependentní	E10.-	insulin-treated non-insulin-dependent diabetes mellitus (disorder)	237599002
hypertenze	Esenciální (primární) hypertenze	I10	essential hypertension (disorder)	59621000
ischemická choroba srdeční	ischemie koronární	I25.9	ischemic heart disease (disorder)	414545008

Tab. 3. Selected attributes of MDMC coded by means of ICD-10 (MKN-10) and SNOMED CT.
(Part III - Current difficulties of a possible cardiological origin).

Attributes of MDMC (in Czech)	Term in MKN-10	MKN-10 code	English equivalent	SNOMED CT (Concept ID)
dušnost	dušnost	R06.8	asthma (disorder)	187687003
bolest na hrudi	bolest hrudníku	R07.4	dull chest pain (finding)	3368006
palpitace	palpitace (srdce)	R00.2	(palpitations) or (awareness of heartbeat) or (fluttering of heart)	161965005
synkopa	synkopa srdeční	R55	syncope (disorder)	271594007
kašel	kašel	R05	cough	158383001
hemoptýza	hemoptýza	R04.2	haemoptysis	158384007

- *Attributes with a too big granularity*, i.e. attributes describing certain characteristics on such a narrow level so that classification systems contain only a term of a more general meaning (e.g. symmetrical pulse of carotids, etc.).
- *Attributes, which cannot be found in classification systems*, e.g. dyslipidemy, etc.

Close cooperation with physicians is essential for solving of such mapping problems. It is often needed to choose the right synonym substituting a certain technical term. It is necessary to do it very carefully not to lose information or not to misinterpret it. In case it is not possible to do it without any lost of information, the better way is to describe a non-coded term by means of a set of several coded terms, possibly with showing mutual semantic relations. If this is not possible, we can polemize with specialists whether these "indescribable" terms (attributes) can be replaced by other more equivalent or more

standard ones. In special cases it is possible to add a certain term to an upcoming new version of a certain coding system. In case it is not possible to use any of the above-mentioned possibilities of solving mapping problems, it is necessary to cope with the fact that mapping will never be 100%. The insufficient mapping process limits the interoperability of heterogeneous systems used for various purposes in healthcare. Restricted interoperability is often inevitable from the very root of the problem, e.g. insufficient harmonization of clinical contents of heterogeneous systems of electronic health records.

6. Conclusion

The analysis of medical reports written in a free text has shown that recording using a free text is very inhomogeneous and not standardized. The biggest problems for computational processing are typing errors, various length of shorten expressions and usage of synonyms. The standardized terminology would bring

benefits to physicians, patients, administrators, software developers and payers. The standardized clinical terminology would help healthcare providers in a way that it could provide complete and easily accessible information that belongs to the process of healthcare (patient's medical record, diseases, treatments, laboratory results, etc.) and it would result in better care of patients.

Despite problems in the usage of international nomenclatures and matathesauri in healthcare in the Czech Republic remain, their use is a necessary first step to enable interoperability of heterogeneous systems of health records. Sufficient semantic interoperability of these systems is the basis for shared care, which leads to efficiency in healthcare, financial savings and reduction of the burden on patients, and therefore in this work we have tried to analyze how the international classification systems could be used best for the needs of Czech healthcare.

Current healthcare information systems enable to collect a variety of clinical information, these systems are linked to clinical knowledge databases, they can search for data, collect data, analyze data, exchange data and they also have a lot of other functions. As the best classification system seems to be SNOMED CT, which can provide a basis for these functions. Information systems can use the concepts, hierarchies and relationships as a common reference point. SNOMED CT may even exceed the direct care of patients. This terminology may, for example, facilitate decision support, statistical processing, monitoring of public health, medical research and cost analysis.

Acknowledgment

The paper was supported by the projects: SVV-2010-265513, 1M06014 project of the Ministry of Education, Youth and Sports CR and by the AV0Z10300504 project of the Institute of Computer Science AS CR.

References

- [1] Bleich H. L., Slack W. V.: Reflections on electronic medical record: When doctor will use them and when they will not, *Int. J. Med. Inform.* 79(2010) 1-4.
- [2] Hoerbst A., Kohl C. D., Knaup P., Ammenwerth E.: Attitudes and behaviors related to the introduction of electronic health records among Austrian and German citizens, *Int. J. Med. Inform.* 79(2010) 81-89.
- [3] Rinner C., Janzek-Hawlat S., Sibinovic S., Duftschmid G.: Semantic Validation of Standard-based Electronic Health Record Documents with W3C XML Schema. *Method Inf Med* 49 (2010), preprint online
- [4] Oemig F., Blobel B.: Semantic Interoperability Adheres to Proper Models and Code Systems: A Detailed Examination of Different Approaches for Score Systems. *Methods Inf Med* 2010 49 2, 148-155
- [5] Lopez D.M., Blobel B.: A development framework for semantic interoperable health information systems. *Int J Med Inform* 2009; 78 (2), 83-103
- [6] Garde S., Knaup P., Hovenga E.J.S., Heard S.: Towards Semantic Interoperability for Electronic Health Records: Domain Knowledge Governance for openEHR Archetypes. *Methods Inf Med* 2007; 46 (3), 332-343
- [7] Nagy M., Hanzlíček P., Přečková P., Kolesa P., Mišúr J., Dioszegi M., Zvárová J.: Building Semantically Interoperable EHR Systems Using International Nomenclatures and Enterprise Programming Technique. In *eHealth: Combining Health Telematics, Telemedicine, Biomedical Engineering and Bioinformatics to the Edge*. Amsterdam: IOS Press, 2008 (Eds. Blobel, B.; Pharow, P.; Zvárová, J.; Lopez, D.) 105-110
- [8] Nagy M., Hanzlíček P., Přečková P., Říha A., Dioszegi M., Seidl L., Zvárová J.: Semantic Interoperability in Czech Healthcare Environment Supported by HL7 version 3. *Methods Inf Med* 2010 49 (2), 186-195
- [9] Zvárová J., Hanzlíček P., Nagy M., Přečková P., Zvára K., Seidl L., Bureš V., Šubrt D., Dostálová T., Seydlová M.: Biomedical Informatics Research for Individualized Life-long Shared Healthcare. In: *Biocybernetics and Biomedical Engineering*, 2009, 29 (2), 31-41
- [10] Přečková P., Špidlen J., Zvárová J.: Usage of the International Nomenclatures and Metathesauruses in Shared Healthcare in the Czech Republic. *Acta Informatica Medica*, 2005 (13), 201-205
- [11] Přečková P., Zvárová J., Špidlen J.: International Nomenclatures in Shared Healthcare in the Czech Republic. In: *Proceedings of 6th Nordic Conference on eHealth and Telemedicine „From Tools to Services“* (Ed.: Doupi P.), 2006, 45-46
- [12] Elkin P. L., Trusko B. E., Koppel R., Speroff T., Mohrer D., Sakji S., Gurewitz I., Tuttle M., Brown S. H.: Secondary Use of Clinical Data. In *Seamless Care Safe Care*. IOS Press, 2010 (Eds. Blobel B., Hvannberg E., Gunnarsdóttir), 14-29
- [13] Stausberg J., Lehmann N., Kaczmarek D., Stein M.: Reability of diagnose coding with ICD-10. *Int. J. Med. Inform.* 2008 (77), 50-57
- [14] IHTSDO: The International Health Terminology Standards Development Organization: SNOMED Clinical Terms® User Guide 2008
- [15] Cornet R.: Definitions and Qualifiers in SNOMED CT. *Methods Inf Med* 2009 (48), 177-183
- [16] Schulz S., Hanser S., Hahn U., Rodgers J.: The Semantics Procedures and Diseases in SNOMED® CT. *Methods Inf Med* 2006 (45), 354-8
- [17] Gault Lora V., Schultz M.: Variations in Medical Subject Headings (MeSH) mapping: from the natural language of patron terms to the controlled vocabulary of mapped lists. *J Med Libr Assoc.* 2002 April; 90(2): 173180
- [18] Khan A. N., Griffith S. P., Moore C., Russell D., Rosario A. C., Jr., Bertolli J.: Standardizing Laboratory Data by Mapping to LOINC. *J Am Med Inform Assoc.* 2006 MayJun; 13(3): 353355
- [19] Louis D. N., Ohgaki H., Wiestler O. D., Cavenee W. K., Burger P.C., Jouvett A., Scheithauer B.W., Kleihues P.: The 2007 WHO Classification of Tumours of the Central Nervous System. *Acta Neuropathol.* 2007 August; 114(2): 97109
- [20] Brierley J.: The evolving TNM cancer staging system: an essential component of cancer care. *CMAJ.* 2006 January 17; 174(2): 155156
- [21] Campbell J.R., Olivek D.E., Shortliffe: UMLS: towards a collaborative approach for solving terminological problems, *J. Am. Med. Inform. Assoc.* 5 (1998), 12-16
- [22] Adášková J., Anger Z., Asche-rmann M., Bencko V., Berka P., Filipovský J., Golář L., Grus T., Grünfeldová H., Haas T., Hanuš P., Hanzlíček P., Holcátová I., Hrach K., Jiroušek R., Kejřová E., Kocmanová D., Kolář J., Kotásek P., Králíková E., Krupařová M., Kyloušková M., Malý M., Mareš R., Matoulek M., Mazura I., Mrázek V., Novotný L., Novotný Z., Pecen L., Peleška J., Prázný M., Pudil P., Rameš J., Rauch J., Reissigová J., Rosolová H., Rousková B., Říha A., Sedlak P., Slámová A., Somol P., Svačina Š., Svátek V., Šabík D., Šimek S., Škvor J., Špidlen J., Štochl J., Tomečková M., Umnerová V., Zvára K., Zvárová J.: A proposal of the Minimal Data Model for Cardiology and the ADAMEK software application (in Czech). Internal research report of the EuroMISE Centre Cardio. Institute of Computer Science AS CR. Prague, October 2002.
- [23] Mareš R., Tomečková M., Peleška J., Hanzlíček P., Zvárová J.: Interface of patient database systems - an example of the application designed for data collection in the framework of Minimal Data Model for Cardiology (in Czech). *Cor et Vasa*, 2002, 44 (4), Suppl., 76
- [24] Eryigit G., Nivre J., Oflazer K.: Dependency Parsing of Turkish. *Computational Linguistics.* 2008, 34(3), 357-389.
- [25] Ringlestetter C., Schulz K. U., Mihov S.: Orthographic Errors in Web Pages: Toward Cleaner Web Corpora. *Computational Linguistics.* 2006, 32(3) 295-340

Contact

Mgr. Petra Přečková

Centre of Biomedical Informatics,
Department of Medical Informatics,
Institute of Computer Science AS CR,
v.v.i.

Pod Vodárenskou věží 2

182 07 Prague 8

Czech Republic

e-mail: preckova@euromise.cz

Matching Medical Websites to Medical Guidelines through Clinical Vocabularies in the Context of Website Quality Assessment

D. Rak¹

¹1st Faculty of Medicine, Charles University in Prague, Czech Republic

Supervisor: Doc. Ing. Vojtěch Svátek, Dr.

Summary

Modern technology offers a wide array of possibilities to publish almost any content freely on the Internet. Because of the importance and delicacy of medical information, the quality of such texts provided to general public seems to be a serious issue nowadays. Unfortunately, the only feasible way to approve the adequacy of the medical information content is human verification. Best practices in medicine are systematically captured by medical guidelines (MGL), which are provided by renowned medical societies and based on results of Evidence-Based Medicine (EBM).

We propose a simple approach exploiting MGL content as a benchmark for the assessment of a content quality in medical web sites (WS). It is based on the idea that the information content or at least the scope of a medical text is reflected in the domain terminology used. We discuss a possible use of this approach in semiautomatic human-based quality verification and various aspects related to its application.

Concept candidates discovered in a MGL and in the tested web pages are matched to UMLS, yielding sets of used medical terms and corresponding concepts. Several aggregation techniques for MGLs were proposed and tested. The two sets are analyzed for overall similarity at term and concept level.

The method was applied on a selected medical topic employing relevant MGL and 100 WS. All the analyzed web pages fell into five distinct categories (corresponding to the target audience). Aggregations for the MGLs were proposed and tested. The average cosine similarity to MGL across all tested WS reached 0.69 whereas the average similarity calculated per each category varied up to 7,6% against the overall number.

The research done is the first step towards automated evaluation of a medical web page content on the basis of MGLs as the quality standard. We describe further tasks which would improve the outputs of comparison and the possibility of its common application.

Keywords: information quality assessment, clinical vocabularies, unified medical language systems (UMLS), evidence-based medicine (EBM), medical guidelines (MGL), information quality, annotation, similarity, concept, content representation

1. Introduction

Modern technology offers a wide array of possibilities to publish almost any content freely on the Internet. There are many widely available methods of creation and publishing of either static or dynamic web pages today. Although insufficiently, the content is at least somehow linked to the creator or publisher in such classical settings. Besides, there is a variety of new techniques commonly called "Web 2.0". This technology brings many further possibilities as it allows the readers of the WS to directly contribute and publish their own texts. It encompasses various systems such as blogs, wiki systems, social networks, discussion groups etc. In this case there is in fact no one accountable for the information content except of the system administrator.

Thanks to powerful tools such as Google [1], the lookup of the information on the Internet based on keyword search is even easier than authoring. Search engines constantly scan and index the space of the Internet without any filtering or censorship. The result of user search is returned in the form of a list of pages sorted by their relevance (wrt. the combination of various criteria managed by the search engine

provider). Even though providers often boast to provide the user with the 'answers', in fact the engine only returns pages that meet the user search the best. However, the sort criteria completely ignore any content verification or filtering of false information, and they do not distinguish certified web pages (that are assumed to be of a high quality).

The only limitation in this information freedom is just the technical skill of the author of the text. However, the lack of knowledge of the problem area and the competence or qualification to speak about the topic is by no means a limitation. This results in a situation when the user looking for certain information may get many inconsistent answers without having the possibility to distinguish between high-quality information, low-quality information, information influenced by an advertisement, or even intentionally misleading information. Because of the importance and delicacy of medical information this problem is perhaps the most striking in this domain. An easy access to a huge amount of information sources in a varying quality (from meta-analyses to general text) for such an important area of life brings problems in many aspects. Correct information can serve to the user very well and bring him/her many positive effects. In global it can also help achieve many savings in the healthcare system. On the other hand relying on misleading and low-quality data may cause a complete opposite effect. The plausibility of discovered information is thus on the very top position between all the quality measures available.

Another sign of widespread easy access to the great amounts of medical information sources is the information overload. It can concern an ordinary user as well as a medical professional. The result might be the omission of very important information for the case in a pile of other unnecessary data. Often mentioned are the hypothetical problems of complicated communication with a patient previously equipped by wrong information.

Unfortunately, the only feasible way to approve the adequacy of the medical information content is human expert verification today. Experts in the field of WS quality assessment usually evaluate the resources in a complex way. Besides semantics they consider many other rather technical features such as the quality and transparency of presentation, presence of contact information or compliance to the web standards [2]. While there do exist some generic standards for these measures, which might be more or less automatically applied to any kind of web pages including medical ones, presently the only feasible way to approve the adequacy and correctness of the medical information content is manual verification [3]. An example of the system which is designed to support expert decision making in quality assessment is AQUA [4] developed within the frame of the MedIEQ project [5]. Semiautomatic content evaluation would be another important improvement of similar tools allowing better efficiency of an expert work. Once assessed the WS are usually provided with the certificate of quality (e.g. HON [6], WMA [7]) or displayed in specialized portals depending on their quality or a topic category.

The other possible way to ensure the high content quality is the situation when the expert him/herself compiles the text about the topic. Such expert-written texts are often provided by renowned medical societies, which warrants a certain level of quality. Apart from the fact that such practice is very expensive, time consuming and thus in fact unusable in a large scale, the big problem still remains unsolved. Even these high quality texts may still become unrecognized between thousands of other available texts. The main present-day challenges for information science in

the area of medical information quality consist in two directions. The first direction consists in the possibility of unambiguous and explicit definition of a unique and consensual version of the truth based on state-of-the-art knowledge. The second challenge is related to the possibility to use this etalon effectively, i.e. find it, compare other documents to it and reference it during the assessment of information quality.

Due to the decentralized creation of new scientific findings, many national specificities occurring in health systems and the existence of a number of organizations aspiring to the position of the highest authority, it is not realistic to expect such a unique and shared version of the truth from any of these entities. The most promising in this context appear to be the activities associated with producing the so-called medical guidelines (MGL) [8]. These documents are systematically prepared and updated by teams of experts and subsequently published under the auspices of prestigious medical societies, medical organizations [9], or agencies specializing in the publication of MGLs [10]. The MGLs are compiled using the principles of Evidence-Based Medicine (EBM) which is based on a hierarchically organized structure of scientific evidence (papers). The aim is to apply primarily the available evidence of the highest strength and significance. The meta-analyses and systematic reviews are on the very top of this hierarchy. MGLs completely cover the area of treatment of the disease in terms of diagnosis, course of the disease, medical procedures, their interchangeability or applicability in different conditions. They even evaluate different methods in relation to their cost or to the difficulties caused to a patient. A very important feature of MGLs is that they are very well structured. Currently, there are already methods aimed to deal with the formalization and with the conversion into an entirely structured electronic version [11]. This can then be implemented for example in hospital systems in combination with electronic healthcare records, or to evaluate information quality of documents on the Internet.

The information quality is defined as the value the information delivers to its user. It

implies that a very important role in the information quality is played by its subjectivity. The very quality of information can be viewed from the four different directions (or dimensions). The first group consists of properties directly related to the essence of the text, e.g. accuracy, objectivity and credibility of information. The second dimension features are setting information into the context of other available information (e.g. completeness, timeliness, relevance or value added). The third dimension is related to properties expressing the adoption of text by a reader; therefore it includes properties such as comprehensibility, ease of understanding, conciseness and logical consistency. The last aspect of information quality is associated with the availability of information to users (e.g. ease of obtaining of the information or its updates or security of access). In order to create some information quality assessment framework, the selection of objective characteristics from the options above needs to be performed in the first place. Based on selected options, information quality metrics are to be created.

The subject of this work refers to the objective characteristics of information quality such as completeness of coverage or lack of coverage of the topic, use of professional terminology, accuracy, reliability, verifiability, and accessibility of information. The subjectivity of information is reflected by the authors of texts as they adjust their texts to particular groups of readers. In the field of medical texts on the Internet it is possible to distinguish between texts intended for general public (adult patients or children) and texts for professionals (e.g. physicians and researchers in medicine). Texts targeted for each of these groups differ in many properties falling into the subjective area. For example, the use of accurate medical terminology enhances the accuracy of expression and is usually very appreciated by the professionals. On the other hand, it may significantly reduce the ease of understanding of the text for the non-professional users.

In the group of subjective characteristics, the influence by the reader category is obvious. However, a similar influence of this categorization may be observed even for the characteristics of more objective nature and thus taking it into account during the assessment of information quality seems to be appropriate as well.

2. Objective

The objective of this paper is to propose a simple approach exploiting MGL content as a benchmark for the assessment of an information quality in medical web sites. Clinical vocabularies are used to discover medical terminology in both groups of texts (MGL and WS). Both sets of terminology are then compared based on extracted data.

The WS content quality will be assessed firstly based on general content match (i.e. based on concepts or topics discovered) and secondly based on similarity of the particular terminology used in MGLs. A partial goal is to propose and evaluate suitable methods of aggregation of terminology in MGLs so that only one single standard for WS quality assessment might be applied in the end. The last goal is to evaluate the overall applicability of this approach in the process of semiautomatic quality assessment. The main focus will be placed on the description of strong and weak aspects of the approach and on the evaluation of its impact on the possibility of practical application.

3. Methods

Existing medical guidelines (MGL) and about one hundred web sites (WS) were looked up for the selected medical topic. For both groups were performed searches for medical terminology. The results obtained from the MGLs were used as a benchmark for evaluating the content of WS.

The topic of "lung cancer screening", where the procedure has been tested, was selected based on following criteria. It needed to be clearly identifiable and delimited, there must have been MGLs available for the topic and the subject needs to be reasonably accessible to the general public. The whole work was done

on English-language papers (MGL, WS) only. The most suitable MGLs were found by the search in existing databases and catalogues of MGLs. In addition to MGL, the search was carried out for existing WS relevant to this issue. Documents used as a corpus were drawn simply as first one hundred links returned by the search engine Google [1] for the search string "lung cancer screening". These WS were subsequently downloaded by the Scrapbook tool [12] and stored locally. The set of documents was manually rounded to one hundred WS after the removal of broken links or sites that were un-downloadable. Similarly excluded were references to the previously selected MGLs, some of which also ranked in the top 100 results returned by the search engine. These MGLs were excluded to avoid bias in comparison with itself. Based on an estimate of the target group of readers (discussed in the introduction), documents were classified into several categories.

Texts of MGLs were annotated using the tools built over the UMLS Metathesaurus [13], [14]. Initially, mapping was performed using the MMTx (MetaMap Transfer) annotator [15], but the use of the Interactive MetaMap tool [16] proved to be more suitable later. Both tools were developed by the NLM [17], an organization that also develops UMLS. In the first phase, the full texts were processed by either of the tools with the list of terms as the output. For each of these terms corresponding concepts were traced by use of SQL querying over the locally stored UMLS database (containing other data sources such as MeSH, ICD-10, etc.). The result of mapping was a list (or a hierarchical tree) of terms and concepts which served as the set of terminology describing the content of MGLs. Similar mapping to UMLS was performed for all of the WS. Mapping products for the two groups were then preliminarily mutually compared.

Unfortunately, the terminology used in the WS often does not match the terminology used in the MGLs and neither the terminology contained in the UMLS. Even though each concept in the UMLS has

assigned a list of synonyms, these terms are again usually scientific terms or names used in other databases of the Metathesaurus. Missing synonyms often comprise colloquial, common, less accurate or abbreviated names of diseases, procedures or medical equipment. These synonyms are necessarily commonly present in texts intended to general public. For instance, the UMLS concept denominated as "Tomography, Spiral Computed" is in reality represented by a range of synonyms, abbreviated or incomplete names and abbreviations such as "CAT Scan, Spiral", "Computed Tomography, Spiral", "Computer-Assisted Tomography, Spiral", "Computerized Tomography, Spiral", "CT Scan, Spiral" and the like. This method was very often referred to as only "scan" in the tested WS, which led to the miss with the UMLS terminology or contrary to ambiguous or incorrect mapping. From the perspective of the document content both sets often seemed to differ, even if it was purely syntactic (terminological), rather than semantic difference. The workflow of the method for the final comparison had to be extended so as to be able to take into account even the terms occurring only in WS (i.e. missing in MGLs or in UMLS).

Set of concepts for the selected topic and corresponding synonymic terms (derived from the first round of mapping), were consequently stored outside of UMLS in a different database structure. All the tested WS (or at least few of the WS) previously annotated by the discovered terminology were then manually checked for an overall coverage of the UMLS terms. If not, the missing terms were added one by one to the stored list of terminology. This adjustment was carried out only for the terms clearly classifiable under the chosen concepts (typically there were variants of existing synonyms). This manual step allowed the subsequent more complete mapping of concepts for all the WS and improved their mutual comparison.

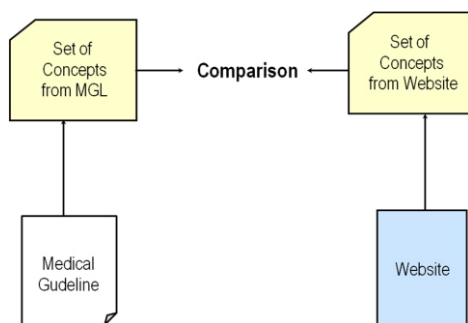


Fig. 1. A diagram showing process of comparison of WS and MGL.

Given that the enriched list of terminology has been stored outside of the UMLS database, the final annotations of WS and MGLs were made using the Super Text Search tool [18], which allows full text searching over the list of documents. Occurring distinct terms and their total number of occurrences in each document were subsequently calculated for annotated files. On the basis of synonymy relations distinct concepts were derived and their occurrences in each of the texts were calculated.

The cosine similarity between the MGLs and WS was calculated from the recorded results. As noted above, there were three suitable MGLs found and used (labeled a_0 , a_1 , a_2). Because of this, it was necessary to decide, which standard was to be the most appropriate to use for WS comparison. In this context, we proposed four different methods of aggregation of sets of mapped terms or concepts respectively.

The four different aggregations were as follows: "intersection" of sets of terms (labeled $a_1 \cap a_2 \cap a_0$) - i.e. distinct terms that were common to all of the MGLs, then "union" ($a_1 \cup a_2 \cup a_0$) - i.e. distinct terms present in at least one MGL, then "sum" ($\text{sum}(a_1, a_2, a_0)$) - i.e. a simply merged list of all terms found, including the number of occurrences, and finally "weighted sum" ($\text{nsum}(a_1, a_2, a_0)$) - in addition, the merger of sets of documents reflects the scope of each MGL (total number of occurrences for all of the terms). The following table demonstrates the process of construction of the four aggregates on fictive data. Evaluation of the suitability of aggregations was performed using the cosine similarity

mutually between all the MGLs and all of the aggregation products. This comparison of mutual similarity was performed first using the full set of all terms found, second based on distinct occurrences of terms present in documents and finally using the mapped concepts.

Comparison of the cosine similarity for all of the WS was made against all the versions of the aggregated standards and against all the MGLs. Comparison was again performed at the level of distinct and total numbers of a medical term which is a number indicating "terminological similarity". Similarly, the similarity was calculated on the basis of distinct (or absolute numbers) of concepts which on the other hand indicates kind of "conceptual similarity". The average cosine similarities of WS against standards were enumerated for each category of documents.

4. Results

The method was applied on a selected medical topic "screening for lung cancer". Test WS were obtained by providing the "lung cancer screening" search string in the Google search engine. Google search returned approximately 2 million records. For instance, similar searches were carried out in the "Yahoo!" search engine [19] (29 million entries) and in the Czech search engine Seznam [20] (only 120 entries). As a corpus of test WS were used the first 100 most relevant results (after adjusting useless links, i.e. broken links and MGLs) returned by Google. Discarded MGLs were positioned in the second half of the top 100 results returned by Google.

All the analyzed web pages fell into five distinct categories (corresponding to the target audience). Aggregations for the MGLs were proposed and tested. The average cosine similarity to MGL across all tested WS reached 0.69 whereas the average similarity calculated per each category varied up to 7.6% against the overall number.

Several MGLs matching the selected topic were found using available information sources, i.e. existing databases or catalogues of MGLs and also freely on the Internet. Three MGLs were chosen and used in the experiment. Discarded MGLs

were either nationally or language-specific, they incompletely covered topic or were much more general in the contrary. In one case it was an older version of one of the three used MGLs.

The tested WS were manually classified into several categories depending on their nature and the target group of users of the text. In case of the chosen topic the WS fell into five categories. The first group consists of pages designed for the professionals in medicine, which were represented by 23 % of the total number of WS. The second important group constituted scientific papers. We have divided this category further into those available in full text or at least in the form of an abstract (23%) and those where there was name or only a very short summary available (5%). An important group of documents were WS targeted for the general public - that is, both for patients (21%) and children (0%). Although the articles for children readers are common for other medical topics (mostly have educational and preventive nature), there were no such documents present in the corpus for the selected topic, probably due to the technical essence of the topic. The last category was created artificially for the texts intended for no particular group of users (28%). This group included namely general reports, statements, newspaper articles and the like.

Tab. 1. Absolute counts of corpus websites in target audience categories.

type	description	count
m	WS for medical professionals	23
p	WS for patients	21
ch	WS for children	0
g	general, news, other	28
mo	scientific papers (full texts or at least abstract)	23
mo/x	scientific papers (restricted access, usually title only)	5

MGLs were annotated by MMTx or the Interactive MetaMap respectively. The results of the process were the texts with mapped scientific terms from UMLS. From the mapped terms it is possible to infer the medical concept which then represent the content of the text. By this procedure 15 distinct concepts relevant to the selected topic area were discovered. Test WS were similarly annotated based on this limited set of medical terminology.

The resulting annotated texts of the two groups (MGLs WS) were then analyzed manually in order to locate gaps in the UMLS mappings. It was found that in MGLs the mapping was almost 100% successful. In contrast, the level of successful mappings in the WS group was estimated to be only about 60%. Both of these findings are fully consistent with expectations and are clearly due to the fact that the UMLS is especially designed to work with texts written by scientific terminology.

All 15 mapped concepts and all their synonyms contained in the UMLS were saved to a new hierarchical database structure. The original list of synonyms was manually expanded in order to include missing terms identified during the review of annotated WS. Subsequently, all the documents were re-annotated by the extended list of terminology reaching a much higher success rate.

In our case, there were three different MGLs available for the comparison. In order to be able to compare the similarity of WS in the future simply against one single standard, one of the goals was to propose some of aggregation techniques for MGLs. Four different versions of the aggregated sets of terms and concepts were compared using mutual cosine similarity.

As the methods of aggregation where the average similarity shows the best match with the initial MGLs were assessed $\text{sum}(a1,a2,a0)$ and $\text{nsum}(a1,a2,a0)$. Similarity of these two methods, depending on the method variant ranged from 0.90 to 0.99. The degree of mutual similarity between sets of terms representing the three MGLs was at a similar level (i.e. approaching 1.00). Aggregation based on an intersection and union resulted relatively less suitable based on the mutual similarity analysis. The lowest average similarities were observed for the intersection.

The analysis of similarities was carried out at three levels of the detail. The first method compared all the terms and reflected the number of occurrences in the text as the weight of the term (in the tables referred to as "similarity (terms)"). The second method also worked with the

terms, but comparison was limited only to their distinct occurrence in the text ("similarity (distinct terms)"). The last method compared the similarity of concepts mapped through the terms found in the text ("similarity (concepts)"). From this standpoint, the highest average similarity values were achieved on the concept and also on the distinct term level. The lowest average similarity was recorded using the absolute number of occurrences of terms. Summary of results for cross-comparison of MGLs and their aggregations are shown in the following tables.

Tab. 2a. Mutual similarity between sets of terms representing each document. Label $a0$, $a1$, $a2$ corresponds to the three MGLs, " $a1 \cap a2 \cap a0$ " corresponds to the intersection, " $a1 \cup a2 \cup a0$ " denotes union, " $\text{sum}(a1,a2,a0)$ " denotes sum and " $\text{nsum}(a1, a2, a0)$ " denotes the weighted sum. Maximum 100% similarity (i.e. identity) is represented by the value 1. The value of 0 indicates absolute dissimilarity of the two sets.

document	similarity (terms)						
	$a1 \cap a2 \cap a0$	$a1 \cup a2 \cup a0$	$\text{nsum}(a1,a2,a0)$	$\text{sum}(a1,a2,a0)$	$a0$	$a1$	$a2$
$a1 \cap a2 \cap a0$	1,00	0,66	0,70	0,71	0,73	0,63	0,69
$a1 \cup a2 \cup a0$	0,66	1,00	0,58	0,61	0,64	0,44	0,64
$\text{nsum}(a1,a2,a0)$	0,70	0,58	1,00	1,00	0,99	0,95	0,95
$\text{sum}(a1,a2,a0)$	0,71	0,61	1,00	1,00	1,00	0,92	0,97
$a0$	0,73	0,64	0,99	1,00	1,00	0,89	0,98
$a1$	0,63	0,44	0,95	0,92	0,89	1,00	0,81
$a2$	0,69	0,64	0,95	0,97	0,98	0,81	1,00

Tab. 2b. Mutual similarity between sets of concepts representing each document. Label $a0$, $a1$, $a2$ corresponds to the three MGLs, " $a1 \cap a2 \cap a0$ " corresponds to the intersection, " $a1 \cup a2 \cup a0$ " denotes union, " $\text{sum}(a1,a2,a0)$ " denotes sum and " $\text{nsum}(a1, a2, a0)$ " denotes the weighted sum. Maximum 100% similarity (i.e. identity) is represented by the value 1. The value of 0 indicates absolute dissimilarity of the two sets.

document	similarity (concepts)						
	$a1 \cap a2 \cap a0$	$a1 \cup a2 \cup a0$	$\text{nsum}(a1,a2,a0)$	$\text{sum}(a1,a2,a0)$	$a0$	$a1$	$a2$
$a1 \cap a2 \cap a0$	1,00	0,91	x	0,91	0,91	1,00	0,91
$a1 \cup a2 \cup a0$	0,91	1,00	x	1,00	1,00	0,91	1,00
$\text{nsum}(a1,a2,a0)$	x	x	x	x	x	x	x
$\text{sum}(a1,a2,a0)$	0,91	1,00	x	1,00	1,00	0,91	1,00
$a0$	0,91	1,00	x	1,00	1,00	0,91	1,00
$a1$	1,00	0,91	x	0,91	0,91	1,00	0,91
$a2$	0,91	1,00	x	1,00	1,00	0,91	1,00

Similarly to the way the sets representing MGLs were mutually compared, the sets

representing WS were compared to MGLs too.

Tab. 2b. Mutual similarity between sets of concepts representing each document. Label $a0$, $a1$, $a2$ corresponds to the three MGLs, " $a1 \cap a2 \cap a0$ " corresponds to the intersection, " $a1 \cup a2 \cup a0$ " denotes union, " $\text{sum}(a1,a2,a0)$ " denotes sum and " $\text{nsum}(a1, a2, a0)$ " denotes the weighted sum. Maximum 100% similarity (i.e. identity) is represented by the value 1. The value of 0 indicates absolute dissimilarity of the two sets.

document	similarity (distinct terms)						
	$a1 \cap a2 \cap a0$	$a1 \cup a2 \cup a0$	$\text{nsum}(a1,a2,a0)$	$\text{sum}(a1,a2,a0)$	$a0$	$a1$	$a2$
$a1 \cap a2 \cap a0$	1,00	0,66	x	0,82	0,71	0,88	0,71
$a1 \cup a2 \cup a0$	0,66	1,00	x	0,96	0,94	0,75	0,94
$\text{nsum}(a1,a2,a0)$	x	x	x	x	x	x	x
$\text{sum}(a1,a2,a0)$	0,82	0,96	x	1,00	0,94	0,86	0,94
$a0$	0,71	0,94	x	0,94	1,00	0,71	0,86
$a1$	0,88	0,75	x	0,86	0,71	1,00	0,71
$a2$	0,71	0,94	x	0,94	0,86	0,71	1,00

This comparison once again took place at three different levels of detail, i.e. at the level of number of terms, at the level of distinct terms and at the level of concepts.

The average similarity across all WS, across all the MGLs (and aggregations) and across all three types of detail reached 0.69. Generally, the lowest similarity was achieved in the analysis at the level of distinct terms (average of 0.56 compared with 0.745 for the concepts and 0.75 for terms). Similarly to mutual comparison of standards, the highest average similarity of corpus WS to MGLs or to their aggregations were found again for the aggregation " $\text{sum}(a1,a2,a0)$ ", respectively " $\text{nsum}(a1,a2,a0)$ ". Slightly lower values were found for the non-aggregated MGLs and the lowest value for the "union" and the "intersection" aggregations.

Average category similarities (quantified by each category of documents) deviated from the overall average in the average range of 6.9% for terms and concepts and in the range of 9% for distinct terms. Generally, the highest correspondence of WS and DP was found for the category "mo" i.e. scientific publications (average 0.78) and the lowest for category "g" (general texts) and "mo/x" (incomplete scientific publications) in contrary.

The difference between comparisons using either terms or concepts is not only technical but also rather semantic. When comparing the sets of terms the resulting number describes “similarity of terminology”. The analysis based on the similarity of concepts is actually a “comparison of the content” of both texts. The average similarity of terminology for the corpus of WS reached 0.76 while the average content similarity reached 0.85.

5. Discussion

This work represented the first attempts to compare content of MGLs and WS. Due to this fact we needed to perform a careful selection of the medical topic in order to be able to demonstrate and verify the process of comparison. The topic had to be chosen so that there existed adequate MGLs (i.e. the topic should be completely covered by MGL and on the other hand it should not form only a subset of MGL). For the chosen topic there were available several MGLs in the end. It allowed us to develop and evaluate some potentially useful ways of representing the MGL content in the form of aggregations of sets of terms. This way a single standard for evaluating the content of WS may be created. During the selection of the topic it was also taken into account whether the first 100 WS evenly represented different groups of readers. Of all the possibly expectable groups the corpus of documents did not represented only the group of WS for children.

The possibility to generalize this approach to any medical issue, however, is associated with many complications.

The first problem is that the procedure anticipates systematic coverage of the whole domain of medicine by MGLs in the future as it relies on it. Today's practice, however, is far away at least in terms of the coverage and organization of creation of MGLs. MGLs creation is a highly distributed process. MGLs are created irregularly and thematically they cover basically just few of the most important areas. MGLs are also linguistically limited to one particular linguistic area, which constitutes another obstacle to their wider distribution, and in their specific application.

The large influence on the applicability of the method also has the coverage of the medical terminology by the UMLS Metathesaurus. Although the UMLS is regularly updated, expanded to more and more new resources and as the result is has very good coverage of concepts, a number of partial terms in the UMLS is still missing. For each concept it offers a range of synonyms. UMLS can thus be used to map the contained terminology to the words and phrases found in specialized texts. The primary objective of UMLS is to be a dictionary of the correct terminology. For this reason there are many missing terms (particularly colloquial, shortened, incorrect or outdated terms), which results in the fact that the mapping often fails for texts written in an everyday language.

These texts use quite a different terminology from those written in a professional language. This has been also shown in this work. The mapping of the

MGLs (written in professional terminology) was almost entirely successful. On the other hand, mapping of WS written in an everyday language achieved success only in 60% of cases.

In order to be able to proceed with the process further and to test the level of the conceptual compliance, we had to extend the list of synonyms manually at one stage. Synonyms were added for all the concepts relating to the selected topic based on discrepancies found in annotated WS. During the manual assessment of WS it proved that the check of the first 10 to 15 papers discloses vast majority of missing terms. The rest of WS were checked just for the sake of completeness. Based on the expanded list of terminology both MGLs and WS were successfully annotated. However, such manual intervention is not generally applicable in bulk for all medical topics and is an obvious weakness of the general application and use of the whole process.

In addition to problems associated with the completeness of the UMLS (i.e. hosting one concept under different names (synonymy)), there are yet other properties of natural languages [21] which pose great obstacles to a reliable term mapping. Probably the most important problem for computer processing of texts is homonymy [22]. For instance: in order to determine precisely which of the meanings of the word is the relevant in a given situation, it is usually necessary to consider the surrounding context and truly understand the meaning of the text.

Tab. 3. Average cosine similarities of WS categories against MGLg (and aggregations). On the vertical axis there are categories of WS and on the horizontal axis there are MGLs (and aggregations) for the three levels of detail (terms, distinct terms and concepts). Red colour shows maximum values while the blue colour minimum values over the different categories and across all data.

category / document	similarity (terms)						similarity (concepts)						similarity (distinct terms)								
	a1 ∩ a2 ∩ a0	a1 ∪ a2 ∪ a0	psum (a1, a2, a0)	sum (a1, a2, a0)	a0	a1	a2	a1 ∩ a2 ∩ a0	a1 ∪ a2 ∪ a0	psum (a1, a2, a0)	sum (a1, a2, a0)	a0	a1	a2	a1 ∩ a2 ∩ a0	a1 ∪ a2 ∪ a0	psum (a1, a2, a0)	sum (a1, a2, a0)	a0	a1	a2
g	0,53	0,43	0,83	0,82	0,81	0,78	0,80	0,75	0,75	-	0,75	0,75	0,75	0,75	0,50	0,53	-	0,56	0,55	0,47	0,53
m	0,59	0,50	0,86	0,87	0,87	0,78	0,87	0,75	0,77	-	0,77	0,77	0,75	0,77	0,51	0,63	-	0,64	0,64	0,51	0,61
mo	0,60	0,53	0,88	0,89	0,89	0,80	0,88	0,72	0,79	-	0,79	0,79	0,72	0,79	0,52	0,67	-	0,67	0,65	0,52	0,66
ch	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
mo/x	0,57	0,48	0,85	0,85	0,86	0,76	0,85	0,64	0,73	-	0,73	0,73	0,64	0,73	0,50	0,53	-	0,57	0,54	0,47	0,56
p	0,56	0,45	0,84	0,83	0,82	0,79	0,81	0,75	0,73	-	0,73	0,73	0,75	0,73	0,52	0,57	-	0,60	0,59	0,49	0,55
all	0,57	0,48	0,85	0,85	0,85	0,79	0,84	0,74	0,76	-	0,76	0,76	0,74	0,76	0,51	0,59	-	0,61	0,60	0,49	0,58

If we leave aside the problems discussed above, one could imagine using this methodology in a semi-automatic process of assessing the quality of Web documents as follows. The first step in this process would be to establish the topic of the tested text using tools such as Aqua [4]. Identified topic would serve both to search for all MGLs relevant to the topic and also to look up all the relevant terminology using UMLS tools. Subsequently, the evaluation of similarity between text and MGLs would be performed and the resulting degree of similarity could serve as an additional basis for decision-making expert.

The document classification into one of the audience categories would help to better interpret the similarity value, or could be considered only for documents belonging to selected categories.

6. Conclusion

The research done is the first step towards automated evaluation of the content of medical web resources using MGLs as a standard of quality. The main goal was to design the process, to evaluate its practical applicability and provide guidelines for further research. At this stage experiments were made on one specific carefully chosen topic for which there existed available appropriate MGLs as well as WS. The topic was elaborated for English-language texts.

In order to obtain a better idea of how to generalize the procedure for any other medical issues, it would be appropriate to make further experiments with randomly selected topics and try to automate the manual steps that process contains. The results show that the similarity between the documents found on the Internet and the MGLs also depends on the category of readers, for whom the text is intended. Category similarities deviated from the overall average in the range of 6.9% (terms and concepts) and 9% (distinct terms) respectively. The highest correspondence of WS and DP was found for scientific publications (average 0.78) while the lowest for category "g" (general texts) and "mo/x" (incomplete scientific publications) in contrary. This categorization of WS was in this work, however, made entirely manually. A promising way to automate this categorization of WS could be the use of the existing functionality of multilingual tool AQUA [2], which was developed for semi-automatic processing of WS. On the other hand, the use of our findings that document classification into a target audience category itself reflects its correspondence with the MGL and thus its quality into some extent may serve directly to expert users of the AQUA system.

Likewise, MGLs search and the selection the best of them was again a purely manual process. In this regard the situation in the future may improve with the way the MGL catalogues on the Internet develop and extend [7], [8]. MGL processing procedu-

res have been designed to enable the automatic aggregation of several available MGLs to one single standard of quality. The best aggregations were selected and tested WS were subsequently compared to this standard.

The next manual step in the procedure was the manual extension of the list of synonyms for UMLS concepts. It was a necessary step for subsequent successful annotation of documents written in an everyday language. On the other hand, it was shown that to find the missing terms it is entirely sufficient to check only the first few annotated WS, ranked according to their search engine relevance.

The set of concepts represents the content of the document only to a certain extent. A great challenge following a simple comparison of sets representing the terminology and content of the documents would be comparing the texts with conflicting claims between the WS and DP [23].

Acknowledgement

Partly supported by MŠMT project no. 1M06014 and by SVV-2010-265513.

References

- [1] <http://www.google.com> (accessed September 10, 2010)
- [2] Curro V., Buonomo P.S., Onesimo R., de RP, Vituzzi A., di Tanna G.L., D'Atri A.: A quality evaluation methodology of health web-pages for non-professionals. *Med Inform Internet Med.* 2004;29(2):95-107.
- [3] Wang Y., Liu Z.: Automatic detecting indicators for quality of health information on the Web. *Int J. Med Inform.* 2006;May 31.
- [4] Stamatakis K., Chandrinis K., Karkaletsis V., Mayer M.A., Gonzales D.V., Labsky D.V., Amigó E., Pöllä M.: AQUA, a system assisting labelling experts assess health web resources. 12th Intern. Symposium for Health Information Management Research (iSHIMR 2007), Sheffield, UK. 2007;18-20 July:75-84.
- [5] Mayer M.A., Karkaletsis V., Stamatakis K., Leis A., Villarroel D., Thomeczek C.: "MedIEQ Quality Labelling of Medical Web Content Using Multilingual Information Extraction." *Stud Health Technol Inform.* 2006;121:183-190.
- [6] <http://www.hon.ch> (accessed September 10, 2010)
- [7] Mayer M.A., Leis A., Sarrias R., Ruiz P.: Web Médica Acreditada Guidelines: reliability and quality of health information on Spanish-Language websites. In: Engelbrecht R et al. (ed.). *Connecting Medical Informatics and Bioinformatics. Proc of MIE2005.* 2005;1287-92.
- [8] Field M.J., Lohr K.N. (Eds):. *Guidelines for clinical practice: from development to use.* Institute of Medicine, Washington, D.C: National Academy Press; 1992.
- [9] <http://www.who.int> (accessed September 10, 2010)
- [10] <http://www.guideline.gov> (accessed September 10, 2010)
- [11] Vesely A., Zvarová J., Peleska J., Buchtela D., Anger Z.: Medical guidelines presentation and comparing with Electronic Health Record. *Int J Med Inform.* 2006;Mar-Apr:75(3-4):240-5. Epub 2005; Sep 15.
- [12] <http://amb.vis.ne.jp/mozilla/scrapbook/> (accessed September 10, 2010)
- [13] Lindberg D.A., Humphreys B.L., McCray A.T.: *The Unified Medical Language System.* *Meth Inform Med.* 1993;32:281-91.
- [14] <http://www.nlm.nih.gov/research/umls> (accessed September 10, 2010)
- [15] <http://mmtx.nlm.nih.gov> (accessed September 10, 2010)
- [16] <http://skr.nlm.nih.gov/interactive/metamap.shtml> (accessed September 10, 2010)
- [17] <http://www.nlm.nih.gov> (accessed September 10, 2010)
- [18] <http://www.galcott.com/ts.htm> (accessed September 10, 2010)
- [19] <http://www.yahoo.com> (accessed September 10, 2010)
- [20] <http://www.seznam.cz/> (accessed September 10, 2010)
- [21] Baud R.H., Ruch P., Gaudinat A., Fabry P., Lovis C., Geissbuhler A.: Coping with the variability of medical terms. *Medinfo.* 2004;11(Pt 1):322-6.
- [22] Rak D., Svatek V., Fidalgo M., Alm O.: Detecting MeSH Keywords and Topics in the Context of Website Quality Assessment. In: *The 1st International Workshop on Describing Medical Web Resources (DRMed 2008)*, held in conjunction with the 21st International Congress of the European Federation for Medical Informatics (MIE 2008), Goteborg, Sweden. 2008: May 27.
- [23] Kaiser K., Miksch S.: Versioning Computer-Interpretable Guidelines: Semi-Automatic Modeling of 'Living Guidelines' Using an Information Extraction Method, *Artificial Intelligence in Medicine (AIM)*, 2008;55-66.

Contact

Mgr. Dušan Rak

U zátíší 545/9

14700 Prague 4–Hodkovičky

Czech Republic

e-mail: dusan.rak@seznam.cz

Speeding Up the Algorithm for Finding Optimal Kernel Bandwidth in Spike Train Analysis

P. Šanda¹

¹Institute of Physiology, Academy of Sciences of the Czech Republic

Supervisor: Doc. RNDr. Petr Lánský, CSc.

Summary

One of the important tasks in the spike train analysis is to estimate the underlying firing rate function. The aim of this article is to improve the time performance of an algorithm which can be used for the estimation.

As there is no unique way how to infer the firing rate function, several different methods have been proposed. A popular method how to estimate this function is the convolution of the spike train with Gaussian kernel with appropriate kernel bandwidth. The definition of what "appropriate" means remains a matter of discussion and a recent paper [1] proposes a method how to exactly compute optimal bandwidth under certain conditions. For large sets of spike train data the elementary version of the algorithm is unfortunately too inefficient in terms of computational time complexity.

We present a refined version of the algorithm which in turn allows us to use the original method even for large data sets.

The achieved performance improvement is demonstrated on a particular results and shows usability of proposed method.

Keywords: action potential, spike train, neural coding, firing rate, convolution, Gaussian kernel, kernel bandwidth, Brent's minimization, parallel computing, MPI

1. Introduction

Many neurophysiological studies are based on the assumption that the majority of information flow between neurons is provided by spikes. Spike trains are believed to form a neuronal code and many coding models successfully predict experimental stimuli features when only the resulting spike train is given. It has been shown that important aspects of the stimuli are coded by the neuron's firing rate, however, the exact procedure how to obtain such a rate from the experimental

data differs from paper to paper and various methods were proposed [2].

Here we consider the method based on the convolution of a spike train with a fixed (Gaussian) kernel, which in result leads to a smooth estimate of firing rate and has been widely used in the past decades [3], [4], [5], [6], [7], [8]. The most difficult part of this method is the selection of the kernel bandwidth, because it affects substantially the "quality" of the estimate, while there is no obvious clue how the optimal bandwidth should be chosen. In [1] authors propose a kernel density estimator based on the mean integrated squared error principle (MISE) and formulate a precise algorithm how to infer optimal (fixed) kernel bandwidth.

For larger sets of recorded spike trains the time complexity of the straightforward version of the algorithm increases, so that it becomes unusable for online queries when studying the features of the method. Here we provide a solution, which improves the time complexity of the implementation. That at the end allows us to work with experimental data in a reasonable manner. It is also worth to note that the proposed solution does not interfere with the actual result of the original method - for the properties and comparison with other methods look in the original paper [1].

2. Methods

2.1 Original method

The firing rate is a non-negative function λ for which the integral $\int_a^b \lambda(t)dt$ gives the expected number of spikes during the time interval $[a, b]$. In the experimental recording we get only one or more trials of spike train data. The problem is how to assess the firing rate $\hat{\lambda}$, which will be as close as possible to the original λ , which is believed to stand in the background of spike discharges. The method is to convolve the

spike train with a specific kernel, thus obtaining a smooth estimate of λ , for example see Fig. 1. In this case we use fixed Gaussian kernel

$$k_w(t) = \frac{1}{\sqrt{2\pi w}} \exp\left(-\frac{t^2}{2w^2}\right)$$

and the problem is reduced to the question how to select the optimal bandwidth w , so that the difference between λ and $\hat{\lambda}$ is minimal. The method itself is beyond the scope of this article, for details see [1]. What is important here is that the core of computation can be summarized in the following statement: find w^* , such that the formula (1) is minimal:

$$C(w) = \frac{1}{n^2} \sum_{i,j} \Psi_w(t_i, t_j) - \frac{2}{n^2} \sum_{i \neq j} k_w(t_i - t_j) \quad (1),$$

where t_i is the time of i -th spike, n is the number of trials,

$$\Psi_w(t_i, t_j) = \int_a^b k_w(t - t_i) k_w(t - t_j) dt, \quad [a, b]$$

defines the time range of the record and k_w is as the kernel used. Since we will study the Gaussian kernel the equation (1) can be rewritten as

$$2\sqrt{\pi} n^2 C(w) = \frac{N}{w} + \frac{2}{w} \sum_{i < j} \left(e^{-\frac{(t_i - t_j)^2}{4w^2}} - 2\sqrt{2} e^{-\frac{(t_i - t_j)^2}{2w^2}} \right) \quad (2),$$

where N is the number of spikes. Note that the term $2\sqrt{\pi} n^2$ is constant and has no effect on w^* . Let us denote the inner term of the sum in (2) as $E(w, t_i, t_j)$. We will denote \mathcal{W} the set of possible values of w , in which we are searching. We denote the size $W = |\mathcal{W}|$ and assume that its points are equidistant.

The straightforward implementation will find the minimum value via evaluating this term in three nested loops:

```
(11) for  $w \in \mathcal{W}$ 
(12) for  $j \in [1, \dots, N]$ 
(13) for  $i \in [1, \dots, j-1]$ 
     $tmp = E(w, t_i, t_j)$ 
    if ( $min > tmp$ )  $min = tmp, w^* = w$ 
```

thus obtaining the time complexity $O(N^2W)$. The selection of \mathcal{W} is dependent on the interval $[a, b]$ and required precision of the optimal value. In a typical case $w^* \ll b - a$ we can select the upper bound of \mathcal{W} to $\log(b - a)$, in the case of bisection (see below) the upper bound is not so vital.

2.2 Bisection

Now we will use that in a typical case where C forms an unimodal function, see Fig 2., though this cannot be asserted in general (such a problem can, for example, occur when searching for bandwidths is smaller than the sampling resolution of input data). Having the unimodal function and a sensible estimate of lower and upper bounds we can use any of the extremum search algorithms based on sectioning the domain. This will reduce loop (11) time complexity from a linear to a logarithmic factor and as a result we obtain the complexity of $O(N^2 \log(W))$. As hinted above while this method helps a lot certain attention needs to be paid before its use.

2.3 Parallelization

Because the evaluation of $E(w, t_i, t_j)$ is independent of the previous computations, it is a natural target for parallelization.

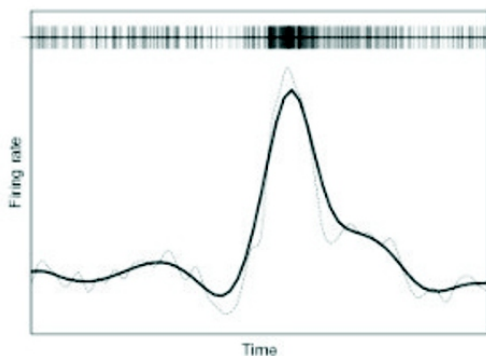


Fig. 1. Illustration of the problem. The thick line is the original λ , the top line shows experimentally measured spike train generated from this function, the thin line is firing rate $\hat{\lambda}$, which we try to optimize.

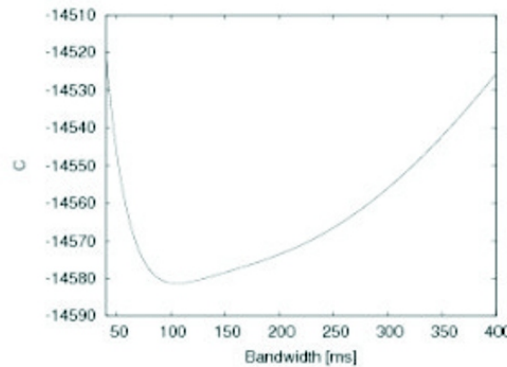


Fig. 2. Typical shape of the function for experimental spike train data.

Splitting the task for the parallel execution at the loop (11) level will not allow us to use parallelization in case of the bisection run, thus we will split the task on p parts at the level (12). That will give us the final estimate for the time complexity of $O(N^2 \log(W)/p)$.

Let us stick with the implementation details now.

2.3.1 Splitting

Since the upper bound in the loop (13) is not constant, trivial splitting of (12) will produce p subtasks $[1, \dots, N/p], [N/p, \dots, 2N/p], \dots, [(p-1)N/p, \dots, N]$ with increasing time complexity of subtasks. At the end this would produce a situation where the first subtasks are completed having the relevant CPUs idling while the last subtasks would still be in computation.

There are more ways how to solve it - (1) move the splitting of task into (13), (2) splitting p tasks in (12) in a proportional way, so that each subtask has the same

computational cost or (3) split (12) into many small subtasks which are successively distributed to CPUs according to their load. In real-life implementation we have chosen (3) because (1) tends to produce high overhead of the parallelization engine and (2) assumes that the underlying CPUs are equivalent in performance and accessibility (that breaks in many distributed environments).

3. Results

3.1 Tuning parameters

The algorithm was implemented based on the sections above, allowing all the strategies - exhaustive search or bisecting both in sequential and parallelized versions. The language used was C++, for parallelization openMPI implementation [9] of MPI standard was chosen, for bisection we used Brent minimization algorithm [10]. In order to find the proper splitting of the subtasks we analyzed measured time demands for a different fragmentation of the tasks, see Fig. 3.

We ran the optimality search for two sets of 1000 and 18000 spikes. In the case of the larger set we could see that taking any value below $f = 500$ gives approximately the same time demands. On the very beginning there is a visible peak caused by the growth of the load by the parallelization maintenance (i.e. the cost of distributing subtasks starts to be larger than computation of subtasks themselves). From $f = 500$ we could see gradual growth caused by the insufficient fragmentation (i.e. some CPUs are needlessly idle and waiting for other unfinished subtasks).

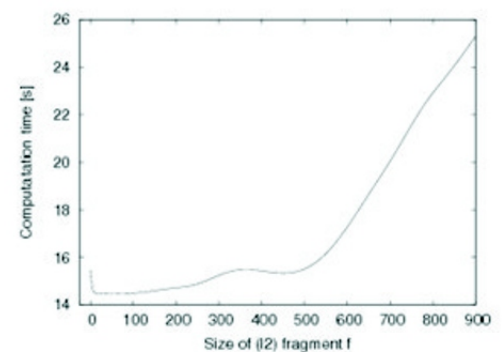
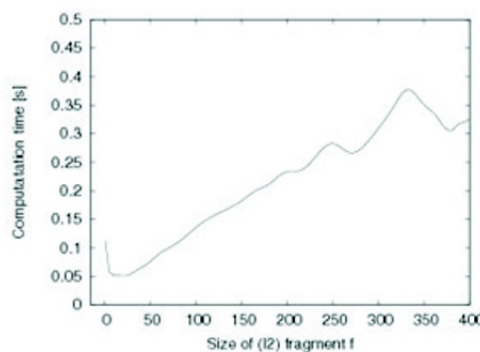


Fig. 3. Figure shows how splitting affects time performance of the computation. The left panel presents the case, where the input data were 1000 spikes, while the input for the right panel was 18000 spikes. Both sets were taken from the real experimental data, bisection was used in this case. f is the size of one (fixed) subtask, that is the number of (13) iterations.

Tab. 1. Comparison of time demands.

Method	Time (min:sec)
Sequential search	58:41
Parallel search	4:14
Sequential bisection	2:43
Parallel bisection	0:09

As the number of spikes in the input set will decrease, this value will also decrease, as the results for the set of 1000 spikes show. Here we could see similar properties as far as the shape is concerned, but the total time needed for computation is now negligible.

This leads to the final choice of $f = 100$, which will be always sufficient for any larger input sets. As it can be seen in the left panel of Fig. 3, it is a reasonable value even for small sets, but that is not so important due to small total time demands. This value is, of course, dependent on the particular computational setting - in our case all tests have been done on a small cluster with 20 CPU cores.

3.2 Real time demands

For the comparison of real-time improvements we offer the table below. The input data and parameters were the same for all the tasks: 18000 spikes, [1;400] ms range for bandwidth, precision of 1 ms ($W=400/1$).

4. Discussion

We have proposed and implemented a parallel algorithm for optimal kernel bandwidth search which has better time performance than its “straightforward” version. Moreover when the function (1) is unimodal on the given range, we can use the bisection version, which reduces the time even more drastically. To check the reasonable ranges, one can do the first trial run which uses only few sampling points (and in fact cannot be omitted even in normal case).

This performance boost does not play an important role in the case of small input sets of spikes, however, it is significant in case of large sets. The whole work was motivated by real demands, when experimental sets of ~20000 spikes were evaluated and, moreover, their subsets also needed to be evaluated the approximate knowledge of the function (1) shape reduced the need for a slow version of the algorithm.

At the end we proposed tuning parameters for an example cluster configuration and provided actual results of the performance improvement.

Acknowledgments

The work was supported by the grant SVV-2010-265 513.

References

- [1] Shimazaki H., Shinomoto S.: Kernel bandwidth optimization in spike rate estimation. *Journal of Computational Neuroscience* 2010; 29: 171182.
- [2] Cunningham J.P., Gilja V., Ryu S.I., Shenoy, K.V.: Methods for estimating neural firing rates, and their application to brain-machine interfaces. *Neural Networks* 2009; 22 (9): 12351246.

- [3] Sanderson A.C.: Adaptive Filtering of Neuronal Spike Train Data. *IEEE Transactions on Biomedical Engineering* 1980; 27 (5): 271 274.
- [4] Richmond B.J., Optican L.M., Podell M., Spitzer H.: Temporal encoding of two-dimensional patterns by single units in primate inferior temporal cortex. I. Response characteristics. *J Neurophysiol* 1987; 57 (1): 132146.
- [5] Richmond B.J., Optican L.M., Spitzer H.: Temporal encoding of twodimensional patterns by single units in primate primary visual cortex. I. Stimulus-response relations. *J Neurophysiol* 1990; 64 (2): 351369.
- [6] Paulin M.G.: Digital filters for firing rate estimation. *Biological cybernetics* 1992; 66 (6): 525531.
- [7] Paulin M.G., Hoffman L.F.: Optimal firing rate estimation. *Neural Networks* 2001; 14 (6-7): 877 881.
- [8] Nawrot M., Aertsen A., Rotter S.: Single-trial estimation of neuronal firing rates: From single-neuron spike trains to population activity. *Journal of Neuroscience Methods* 1999; 94 (1): 81 92.
- [9] Gabriel E., Fagg G.E., Bosilca G., Angskun T., Dongarra J.J., Squyres J.M. et al.: Open MPI: Goals, Concept, and Design of a Next Generation MPI Implementation. In: *Proceedings, 11th European PVM/MPI Users' Group Meeting* 2004; 97104.
- [10] Brent R.P.: *Algorithms for minimization without derivatives*. Dover Pubns; 2002.

Contact

Mgr. Pavel Šanda

Institute of Physiology,
Academy of Sciences of the Czech Republic
Václavská 1083
142 20 Prague 4
Czech Republic
e-mail: ps@ucw.cz

Simple Method of Economical Analysis of Diagnosis Procedure (Used in Screening of Thyroid Gland Diseases in Pregnant Women)

Z. Telička¹, J. Jiskra¹, D. Springer², Z. Límanová¹

¹Third Department of Medicine, General University Hospital and First Faculty of Medicine, Charles University in Prague, Czech Republic

²Department of Clinical Biochemistry and Laboratory Medicine, General Teaching Hospital, First Faculty of Medicine Charles University

Supervisor: Prof. MUDr. Štěpán Svačina, DrSc., MBA

Summary

Thyroid gland diseases are frequent in pregnant women. In the Czech Republic, there is not a screening for this disease. Our objective was to evaluate the initial cost of the study of screening of thyroid gland disease in pregnant women.

Methods: We used simple statistical methods for comparing the cost of the study (examination of TSH and TPOAb) with the costs and efficiency when TSH-only or TPOAb-only screening would be performed.

Screening of TSH only is 2.7 times cost-effective than screening of TPOAb. Screening of TSH only may lead to miss 77 patients with TPOAb positive results and screening of TPOAb only may lead to miss 77 patients with TSH positive results.

Discussion: To achieve better sensitivity, it is necessary to perform screening of TSH and TPOAb together.

Keywords: thyroid gland disease, thyro-pathy, pregnancy, cost-effectiveness, screening, iq, qaly

1. Introduction

Thyroid gland diseases are frequent in pregnant women [1]. The untreated dysfunction of the thyroid gland during the pregnancy may have a negative impact both to a mother and mostly to a child. It can cause low concentrations of thyroid gland hormones which are necessary for the proper development of the child brain, especially during the beginning of pregnancy [2]. Screening tests for thyroid disease consist of measurements of thyroid peroxidase antibodies (TPOAb), thyroid stimulating hormone (TSH) levels and levels of thyroid hormones (T3 and T4) in the blood. Anti-TPO antibodies indicate an autoimmune reaction in which the body sends immune cells to attack the thyroid gland. In foreign studies, about 11 % of

them have proven the disease after a blood examination [1]. Observational studies have concluded that the treatment of maternal hypothyroidism with the thyroid hormone reduces the incidence of gestational hypertension in the mother [3] and improves IQ in the child [4], [5]. Positive TSH results (and negative TPOAb) may lead to improper development of child's brain during the pregnancy and positive TPOAb results (and negative TSH) may lead to the post partum thyroid disease. High ratio of incidence makes clinicians to introduce a systematic screening for this disease. Currently, there is an ongoing discussion whether to establish the screening targeted to all pregnant women or to women with a risk factor and what parameters to examine. All these factors are influenced by the economical aspect from two sides:

- 1) Is the cost for screening adequate to a number of positive matched examinations?
- 2) What benefits does it bring to a mother and a child the successfully treated thyroid gland disease after the positively matched examination?

1.1 Already performed foreign study

The cost effectiveness analysis was already performed on screening of pregnant women in USA. Three strategies were used on the tested group: Women with no screening; one-time screening of TSH and one-time screening of TPOAb. Abnormal screening tests were followed by further testing and hypothyroid women were treated by thyroxine until the test results came to the normal level. Than the cost of the effectiveness analysis was performed, it included the cost of:

- visits at clinic and endocrinologist,
- laboratory screening tests for TSH or TPOAb,
- levothyroxine treatment.

In the study the cost of a low IQ was estimated. Authors of the project worked with consideration, that there is a significant fall of child's IQ if a mother is not treated during the early state of pregnancy. A one-point difference in IQ is associated with about a 2% difference in earnings [7]. With this assumption they found a loss of USD 15 513 in future earnings per IQ point. Based on the follow-up to the Haddow study [8] they assumed that children of untreated overtly and subclinically hypothyroid mothers have on average a 9.2- and 4.2-point decrement in IQ (after adjusting by 0.8 point for differences in socioeconomic status compared with controls). By developing the Markov Model [9] it was possible to find out the Incremental cost-effectiveness ratio for the child per quality adjusted life year (QALY) if the woman is treated for the thyroid gland disease during the pregnancy compared to [6]. Screening pregnant women in the first trimester with TSH was cost-saving compared with no screening, while screening with TPOAb was cost-effective compared with TSH screening at USD 15 182 / QALY.

1.2 Initial study performed in the Czech Republic

A multicentre pilot study has started in 13 regions of the Czech Republic supported by the General Health Insurance Company focused on examinations of thyroid gland diseases in pregnant women. The aim of this study was to find out initial data which could help the clinicians and the insurance company with the decision, whether the screening of women could be beneficial to women and cost-effective for the insurance system.

In our study we tried to find out the costs of this initial study compared to positively-screened tests and simulate the costs in the current situation when the screening is not paid by the insurances companies. We also simulated the costs in the situation when the screening could contain only the TSH or only the TPOAb serum tests.

2. Methods

During the 11-13 week of pregnancy, pregnant women were tested for TSH and TPOAb. If the laboratory reported positive results, women were recommended to visit the endocrinologist who eventually started the treatment. Current costs in the insurance system for testing of these three results in the standard test cost 518 CZK (TSH 140 CZK; TPOAb 378 CZK). Since there it is not a standardized method for TSH and TPOAb, each laboratory has developed its own test methods and ranges. In our study we used normal levels used in each laboratory. Orientation normal levels for TSH are 0.1-3.7 mU/l and for TPOAb ≤ 34 IU/ml. From the study all the patients positively screened due to the suppressed TSH were removed to exclude healthy women with physiologic TSH suppression in the first trimester of pregnancy. Therefore we analyzed only the patients with elevated TSH (>3.7 mU/l). We calculated the summary cost for the initial screening and according to the number of positive results of tested parameters we specified the cost of one positive diagnosis.

3. Results

The costs for TSH and TPOAb screening was 1 373 218 CZK. 612 (23 %) patients were tested positive in one of these two parameters. It means the cost for one positive result of 2 243 CZK.

We tested if it would be sufficient to test women for TSH or TPOAb only. Positively tested (up of reference range) for TSH were found 269 women (10.14 %) and from this 77 (28.63 %) where TPOAb positive and 192 (71.37 %) were TPOAb negative. The cost for this screening would be 371 140 CZK; 1 380 CZK for one TSH-positive screening. But 2382 (89.85 %) TSH-screening negative women have 143 (6 %) TPOAb positive results. It means that TSH-only screening may lead to miss 77 patients with TPOAb positive results.

Positively tested for TPOAb-screening only were 220 (8.3 %) women and from this 77 (35 %) have TSH positive results. The cost would be 1 002 078 CZK; 4 555 CZK for one TPOAb-positive screening. But 2 431 (91.70 %) TPOAb negative women have from this 192 (7.9 %) TSH positive screening results. It means that TPOAb-only screening may lead to miss 77 patients with TSH positive results.

4. Discussion

From the results we have found out that both TSH and TPOAb screening tests are not substitutable and they have to be screened together although the costs for the TSH screening is 2.7 times most cost-effective than TPOAb screening. Screening of one of the above parameters could mean lost of 77 positive women.

Additionally, screening of these two parameters is important due to fact that positive results of each parameter mean for the mother and child a different risk: Positive TSH can cause improper development of child's brain during the pregnancy and positive TPOAb results can cause the post partum thyroid disease.

Acknowledgements

This work was supported by a grant from the Czech Health Ministry IGA NS10595-3 and by the project SVV-2010-265513.

References

- [1] Hollowell J.G., Staehling N.W., Flanders W.D., Hannon W.H., Gunter E.W., Spencer C.A., Braverman L.E.: Serum TSH, T(4), and thyroid antibodies in the United States population (1988 to 1994): National Health and Nutrition Examination Survey (NHANES III). *J Clin Endocrinol Metab.* 2002 Feb;87(2):489-99.
- [2] Springer D., Zima T., Limanova Z.: Reference intervals in evaluation of maternal thyroid function during the first trimester of pregnancy. *Eur J Endocrinol.* 2009 May;160(5):791-7. Epub 2009 Feb 19.
- [3] Leung A.S., Millar L.K., Koonings P.P., Montoro M., Mestman J.H.: Perinatal outcome in hypothyroid pregnancies. *Obstet Gynecol.* 1993 Mar;81(3):349-53.
- [4] Haddow J.E., Palomaki G.E., Allan W.C., Williams J.R., Knight G.J., Gagnon J., O'Heir C.E., Mitchell M.L., Hermos R.J., Waisbren S.E., Faix J.D., Klein R.Z.: Maternal thyroid deficiency during pregnancy and subsequent neuropsychological development of the child. *N Engl J Med.* 1999 Aug 19;341(8):549-55.

- [5] Man E.B., Brown J.F., Serunian S.A.: Maternal hypothyroxinemia: psychoneurological deficits of progeny. *Ann Clin Lab Sci.* 1991 Jul-Aug;21(4):227-39.
- [6] Dosiou C., Sanders G.D., Araki S.S., Crapo L.M.: Screening pregnant women for autoimmune thyroid disease: a cost-effectiveness analysis. *Eur J Endocrinol.* 2008 Jun;158(6):841-51.
- [7] Grosse S.D., Matte T.D., Schwartz J., Jackson R.J.: Economic gains resulting from the reduction in children's exposure to lead in the United States. *Environ Health Perspect.* 2002 Jun;110(6):563-9.
- [8] Klein R.Z., Sargent J.D., Larsen P.R., Waisbren S.E., Haddow J.E., Mitchell M.L.: Relation of severity of maternal hypothyroidism to cognitive development of offspring. *J Med Screen.* 2001;8(1):18-20.
- [9] Sonnenberg F.A., Beck J.R.: Markov models in medical decision making: a practical guide. *Med Decis Making.* 1993 Oct-Dec;13(4):322-38.

Contact

Mgr. Zdeněk Telička

Third Department of Medicine,
General University Hospital and First
Faculty of Medicine, Charles University
Prague,
U Nemocnice 1
128 08 Prague 2
Czech Republic
e-mail: zdenek@telicka.cz

Identification of Units and Other Terms in Czech Medical Records

K. Zvára, V. Kašpar¹

¹Center of Biomedical Informatics, Institute of Computer Science, Academy of Sciences of the Czech Republic

Supervisor: Doc. Ing. Vojtěch Svátek, Dr.

Summary

Healthcare documentation in the Czech Republic usually has the form of a free text formatted just using spaces, tabs and line breaks. Extracting information from such a documentation is a challenge that if fulfilled would allow to use Czech medical reports by physicians with no knowledge of the Czech language as well as information transfer to a structured form. It is possible to approach this task as a task of finite-state machine, as a task of the linguistic analysis or as a task of statistics. This article summarizes our findings gained using finite-state machines and using commonly used code lists. Excerpts from real medical reports are translated to English in a way that demonstrates the same or similar problems as in the Czech language. Original Czech excerpts are available in the Czech version of this article.

Keywords: natural language processing, healthcare documentation, medical reports, EHR, finite-state machine, regular expression

1. Czech Healthcare Documentation and Medical Reports

Czech healthcare documentation is usually in the form of a free text. The content of healthcare documentation is regulated by a Ministry of Health ordinance. Legal aspects of healthcare documentation are clarified by e.g. JUDr. MUDr. Roman Žďárek, PhD. in [1]. In the year 2001 the law No. 260/2001 Sb. amending law No. 20/1966 Sb, "on public healthcare", added new paragraphs 67a to 67d regulating healthcare documentation and the National Healthcare Information System (NHIS). This law explicitly states that keeping healthcare documentation as well as its other treatments is considered as processing of personal data according to the law No. 101/2000 "on personal data protection".

According to the law on public healthcare, § 67b section 4, healthcare documentation contains:

- personal data of a patient in the extent necessary for personal identification and history,
- information on a disease, on case history and results of examination and treatment and on other important conditions associated with the health state of a patient and with healthcare delivery procedures.

According to the law on public healthcare, § 67b section 4, records in healthcare documentation must be conclusive, true and perceivable; it should be continuously updated; it must be dated, the author must be identified and it must be signed by its author. All corrections in healthcare documentation must be done by adding new dated entries. Such entries must contain author identifications and authors' signatures. The original record must remain readable.

Appendix 1 of the ordinance No. 385/2006, "on healthcare documentation", regulates the minimal content of healthcare documentation containing information on the examination and the treatment provided. This part called medical report must contain at least: a) information on patient's health state including results of laboratory and other examinations, b) information on treatment provided and patient response, c) recommendations to further healthcare delivery.

Czech medical reports usually bear the form of an unstructured text. The fact that physicians are organized in a professional organization and share a common education implies that medical reports created by different physicians are organized similarly. Especially the order of medical report segments is very common.

The text is usually formatted just using spaces, tabs and line breaks. No markup is commonly used. New medical reports are either created from scratch (starting with an empty text file), from a template (including automatically generated content including patient identification, health insurance agency, physician's name and today's date) or by copying and modifying the last created patient's medical report. The cost of time spent creating the medical reports is usually significantly reduced in cases when reports are created using templates and/or modifying older reports. Therefore creating reports from scratch is not very common. Using old medical reports as templates bears the risk of leaving no longer valid information in the new report. Such problems connected with copying and updating an old report content have been observed also in other countries as described in details e.g. in [2].

Another specificity of Czech medical reports is occasionally inserted a content copied from another sources. Such a content usually originates from a machine of a cooperating unit (e.g. biochemical laboratory). Such an inserted content is sometimes formatted in the form of textual-table, where formatting is retained in case of fixed-width fonts. Data of biochemical examinations are mostly stored in rows, where each cell represents a single result. The common content of such a biochemical analysis data row includes "test name", "detected value", lower and upper limits and a graphical representation of a relation among a lower limit, detected value and upper limit.

Czech medical reports contain large number of typing errors, acronyms and abbreviated words, see e.g. [3]. Used abbreviations usually have a different meaning in different contexts. Punctuation and spacing is also often wrong.

Disorder names, diagnoses and other words are commonly abbreviated, e.g. "diabetes mellitus" is sometimes abbreviated to "diab.mel." or to "DM". Data of history are usually not a plain list of diagnoses and other events but often encoded in linguistic properties of a sentence. The therapeutic segment of medical reports sometimes also bears parts meant as informative for the patient, e.g. "should pat. have daily soup and also main meal, then pollard to soup here in Biocenter".

Besides the term "healthcare documentation" we commonly hear the term "health documentation". We could not find the term "health documentation" in any official document of the Czech Ministry of Health. But otherwise we encounter this term very often, especially on the Internet. We feel that the term "health documentation" covers only a part of information commonly found in "healthcare documentation", specifically the information on the subject of health state of the patient. This interpretation of the term "health documentation" would conform to tools for data collection on patient's health state using the electronic health record (EHR).

2. The Purpose of Identification

Free-text medical reports probably suit the needs of physicians who speak the same language to provide health care services to patients. Using the information stored in free-text medical reports is constrained by physicians' interpretation. Current usage of such information requires manual input for the reason of the interpretation requirement.

Physicians not speaking the same language as the physician who produced the medical report are currently not able to fully utilize a medical current free-text report in Czech. To use it they need to get it translated by a physician speaking both languages of the medical report provider and medical report user. Often physicians can use only information they suppose they understand (e.g. latin-named diagnoses, ICD-10 codes, pharmaceuticals, commonly-formatted values like blood pressure). Therefore one of practical usages of the information extraction methods is a translation of medical reports to another language.

Scientific and standardization organizations engaged in medical informatics issues are specifying standards to retain medical information in classified objects. Such standards include HL7 CDA and EN 13606 (EHRcom). Hence another practical usage of information extraction is converting free-text medical reports to a structured form by either marking the original text (HL7 CDA) or by transferring information to a standalone graph (EHRcom).

Information from healthcare documentation gathered in medical reports may be used for other purposes. It may serve consolidation of data gathering to combine the creation of a medical report and the creation of the bill for healthcare provided to a health insurance company, to a patient etc. Similarly, it can support systems like processing medical guidelines that can use information from medical reports to just-in-time assist physicians and e.g. provide a warning in the case when the physicians decides in contradiction with installed processing guidelines. It can be seen that the data from healthcare documentation may be used for a variety of other uses. The detailed description of other possible uses is described in [4].

All three main purposes of information extraction from medical reports may be achieved by extraction into a structured health record with a graph structure. Very important seems the extraction of information needed for emergency care.

Main information for emergency care is

- the list of known allergies that is usually a part of personal history,
- the list of current and past diagnoses including the last blood pressure measurement and biochemical laboratory results,
- patient identification, especially name, surname and healthcare insurance number.

These needs may be easily satisfied by the structure provided by EN 13606, International Classification of Diseases (ICD-10) and selected codes from LOINC or NČLP.

3. Information Extraction from Medical Reports

Information extraction from Czech medical reports using non-statistical methods was a subject of a diploma thesis of Jiří Semecký [5]. He provided two definitions of the free-text medical report semantic analysis. The first one defines it as finding the algorithm that finds fragments of the medical report that have sense defined in the knowledge base. The second one defines it as finding the algorithm for the semantic analysis and markup of segments found in the medical report. In the thesis [5] arguments why the linguistic analysis is not suitable for free-text medical reports are brought in. These arguments include the fact that free-text medical reports usually do not contain whole well-formed sentences, information is rather stored in keywords. The second methods of free-text records cited in [5] is finite-state machine analysis, specifically using regular expressions based rather on the format than on a particular language.

Peter Smatana [6] referring to [5] continues in experiments with the linguistic and semantic analysis. Similarly to [5] neither in [6] are used any codebooks of medical terms nor any general vocabularies. Extending the finite-state machine analysis (using regular expressions) by the linguistic analysis the thesis [6] gets slightly better results. Though the thesis [6] is written in the Slovak language there were analyzed Czech medical reports.

Frequent use of abbreviations in Czech free-text medical reports seems to be a serious hurdle of automatic identification of named entities in Czech medical texts. Problems of abbreviations have been studied and described by various authors, e.g. [7], [8]. They usually proposed creation of abbreviation dictionaries with the possible assistance of the linguistic analysis. Tsong O. Cheng brought in an interesting statement in "Words to Editor" in the Journal of Royal Society of Medicine [7]. He stated that more than 90 % of junior physicians do not understand abbreviated words in medical reports outside their speciality.

Although his contribution is outside the field of natural language processing it clearly shows that extensive abbreviation use is a serious problem because the meaning of individual abbreviation depends on its context.

The following part contains information on our approach to identification of numeric values and terms. Proposed procedures have been evaluated on 107 medical records.

4. Identification of Numeric Values

The procedure of identifying is shown in Figure 1. The first part of the processing is based on input text adjustment, the second part uses finite automata to recognize defined structures.

The adjustment of text consists especially in text tokenization. The tokenization process provides us with a list of tokens to be used in the second part of the processing.

For example the text „Krevní tlak:145/90“ is transformed to the following list of tokens (containers), see Table 1.

Individual finite automata are implemented as classes sharing a common interface. We denote these classes as "recognizers". Recognizers use one or more subsequent containers. Containers contain information on the underlying source. Currently we are using these types of containers:

Date

The container contains recognized date. Date is recognized by the recognizer-class *ManualRecognizer*. Subsequent containers are regarded as recognized date when they contain proper separators and information on date (day, month, year). Dates are recognized from containers of the type *DelimitedNumber*.

DelimitedNumber

The contains recognized numbers separated by the same separator represented

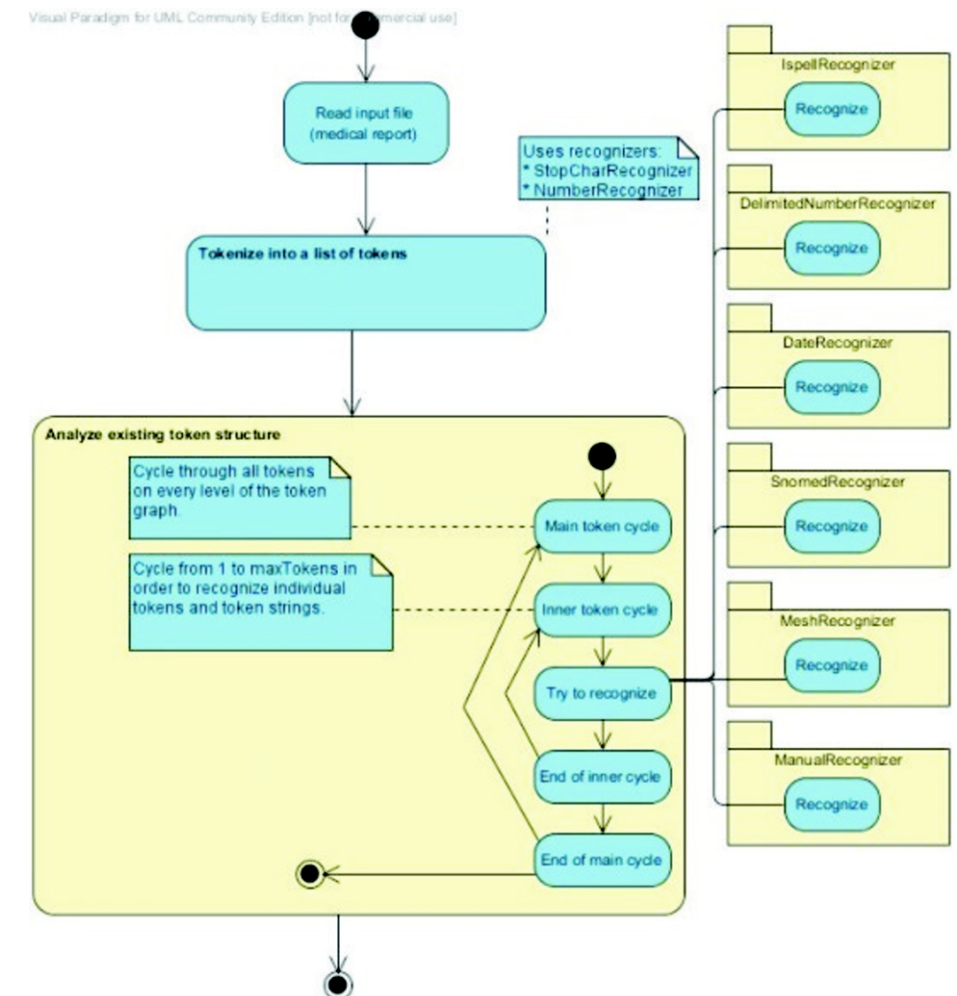


Fig. 1. Activity Diagram of the Experimental Algorithm.

by the container type *SpecialChar*. By this way Czech identification numbers and values of blood pressure measurement are recognized.

Dictionary

This contains recognized Czech words. This also contains a word class. This container type is created by the *IsPELLRecognizer* based on the modified Czech IsPELL dictionary. The modification was needed to identify the word class.

MedicalTerm

This contains recognized items of SNOMED CT and MeSH. This container type is created by the *SnomedRecognizer*

and *Mesh-Recognizer*. These recognizers work on top of basic containers (unrecognized text) and *Dictionary* containers.

We are able to identify the patient insurance number and blood pressure. The Czech patient insurance number contains the date of birth, information on sex and the majority of identifiers (assigned in the last 40 years) also contain validity verification information. We were able to identify patient insurance number in 85 % of reports. We were not able to automatically identify the patient identification number in these cases:

- the report did not contain the patient insurance number (contrary to regulation on healthcare documentation) or
- the record contained only the date of birth, in some cases only the year of birth or

Tab. 1. List of tokens representing „Krevní tlak: 145/90“.

Container	SpecialChar	Container	SpecialChar	Number	SpecialChar	Number
„Krevní“	„ “ (1x)	„tlak“	„:“ (1x)	„145“	„/“ (1x)	„90“

- the patient insurance number (personal identification number) had an invalid format (skipped zero).

The structure corresponding to blood pressure measurement has been successfully recognized in 96 % of reports. In other cases the blood pressure values were not recognized because the blood pressure was not written in the form of SBP/DBP but separately. We have used checks like separator type ("/") maximal acceptable values of systolic blood pressure and the rule that values of systolic blood pressure are always higher than corresponding values of diastolic blood pressure. In one exceptional case the recognized value did not correspond to blood pressure but to parameters of the ECG result.

6. Term Identification

To identify medical terms we have used the SNOMED CT classification system in the version of 2008 [9], International classification of diseases (ICD-10) in the Czech version distributed by ÚZIS [10], codebooks of NČLP (Czech national medical classification system) [11] and the Czech version of MeSH [12].

The most successful identification of medical terms was done by the Czech version of MeSH.

Contrary to the Czech version of ICD-10 there are nearly no abbreviations. Summarized information on processed medical reports are given in Table 2. The words there are in this case marked by tokens detected by tokenization. Among words there are also single numbers (e.g. information on blood pressure "145/90" contains two tokens recognized as words). The first ten the most often recognized Czech descriptors of the MeSH classification are displayed in Table 3.

We have tried to identify terms by matching text chunks (up to five successive words) with codebook values. The searching algorithm always tried to find the exact match first and if the exact match has not been found, it continued by searching for a partial match with a fixed chunk start. In case if chunk contained a period character ('.'), the algorithm replaced that period character with the wildcard (any content). In spite of that the hit rate of identification by SNOMED CT (English version) and by ICD-10 were very low.

SNOMED CT was helpful when finding terms that have the same name in English and in Czech.

Such terms are usually Latin, e.g. "diabetes mellitus" (SNOMED CT code 73211009). Similarly SNOMED CT was helpful in identifying physical units (e.g. "mmHg", SNOMED CT code 259018001) and some names of pharmaceuticals from the "therapy" segment of medical reports). We were not expecting success in identification, simply because the Czech version of SNOMED CT does not exist and medical terms vary according to a language.

We were very disappointed by a low hit rate of the Czech version of ICD-10 published by ÚZIS. The Czech version of ICD-10 is a completely translated original ICD-10 codebook. Items are not named in full. Like in medical reports names in Czech ICD-10 extensively use abbreviations and acronyms. Above mentioned "diabetes mellitus" is sometimes shortened to simple "DM". Many items use more than one abbreviation. The name of the T449 item is (loosely translated) "Ped. x nonmot.veh.; non-trf.a.;comm.", the name of the V0920 item is (loosely translated) "Ped. x injured during trf.acc. o. and uns. mot. v.".

Tab. 2. Summarized information on processed medical reports.

Number of medical reports	107
Number of words (tokens)	66.376
Number of recognized tokens (Mesh CZ)	1.126
Average number of recognized tokens in message	10,52
Recognized tokens / total tokens	1,7 %

Tab. 3. Ten the most often recognized descriptors of the MeSH classification.

Code	English descriptor	Occurences
H01.671.691	pressure	63
A07.541	heart	62
C14.907.489	hypertension	39
A01.047	abdomen	39
C14.280.067	arrhythmia	37
A01.456	head	34
C23.550.288	disease	33
A01.598	neck	30
D04.808.247.222.284	cholesterol	29
E01.370.370.380.650	pulse	28

These two examples show that "vehicle" is sometimes abbreviated to "veh." and sometimes to "v.". Because free-text medical reports themselves contain extensive numbers of abbreviations, the current Czech ICD-10 is practically unusable for the task of automatic information extraction from free-text Czech medical reports.

6. Conclusions

Currently there is no other codebook than MeSH [12] that contains Czech terms of diagnoses that could be used for their identification in medical reports.

Codebooks of NČLP [11] do not contain list of diagnoses so they cannot be used for such a basic task. The codebook of units of NČLP could be useful to at least identify units. Unfortunately NČLP codebooks are neither marked by registered OID nor connected to UMLS [13].

As far as the Czech version of SNOMED CT would exist, it might be easier to identify and classify medical terms. A candidate for this task is also the Czech version of MeSH published by the Czech Medical Library. MeSH is indexed by UMLS and therefore its codes may be mapped to other classification systems, e.g. SNOMED CT, ICD-10, LOINC. Named entities mapped to such classification systems like SNOMED CT would be usable in structures defined by the HL7 standard or EHRcom.

This article describes the current status of our work. We will continue to use non-statistical methods to study the possibility of extracting information from Czech medical reports.

Acknowledgement

The paper has been partially supported by the project 1M06014 of the Ministry of Health, Youth and Sports CR and by the project SVV-2010-265513

References

- [1] Ždarek R.: Vedení zdravotnické dokumentace a její náležitosti. Zdravotnické noviny 30.6.2009
- [2] Hammond K., Helbig S., Benson C., Brathwaite-Sketoe B.: Are Electronic Medical Records Trustworthy? Observations on Copying, Pasting and Duplication. AMIA Annual Symposium Proceedings 2003; 269273.
- [3] Přečková P., Zvárová J.: The Role of International Nomenclatures and Standards in Travel Shared Health Care. Mihalas, G.; Saka, O.; Mazzoleni, T.; Blobel, B.; Pharow, P.; Gülkesen, K. eds. Travel Health Informatics and Telehealth. European Notes in Medical Informatics 5, 2009, 162-164
- [4] Safran Ch., Bloomrosen M., Hammond W., Labkoff S., Markel-Fox S., Tang P., Detmer D.: Toward a National framework for the Secondary Use of Health Data: An American Medical Informatics Association White Paper. Journal American Medical Informatics Association 14 (1):2007, 1-9
- [5] Semecký J., Zvárová J.: Multi-mediální elektronický záznam o nemocném v kardiologii, Univerzita Karlova v Praze, Matematicko-fyzikální fakulta, Praha, 2001
- [6] Smatana P., Paralič J.: Spracovanie lekárskech správ pre účely analýzy a dolovania v textoch, Technická univerzita v Košiciach, Košice, 2005
- [7] HaCohen-Kerner Y., Kass A., Peretz A.: Baseline methods for automatic disambiguation of abbreviations in Jewish law documents. Lecture Notes in Computer Science, Advances in Natural Language Processing Baseline methods for automatic disambiguation of abbreviations in Jewish law documents Vol 3230 Springer 2004, 58-69
- [8] Yu H., Hripcsak G., Friedman C.: Mapping Abbreviations to Full Forms in Biomedical Articles. , Journal of the American Medical Informatics Association 9 no.3, 2002, 262272
- [9] Systematized Nomenclature in Medicine – Clinical Terms, International Health Terminology Standards Development Organisation
- [10] Mezinárodní klasifikace nemocí, Ústav zdravotnických informací a statistiky / World Health Organization
- [11] Národní číselník laboratorních položek, Ministerstvo zdravotnictví ČR, <http://ciselniky.dasta.mzcr.cz>
- [12] Medical Subject Headings, Národní lékařská knihovna / United States National Library of Medicine
- [13] Unified Medical Language System, United States National Library of Medicine

Contact

Ing. Karel Zvára

Center of Biomedical Informatics,
Institute of Computer Science,
Academy of Sciences of the Czech Republic
Pod Vodárenskou věží 2
182 07 Prague 8
Czech Republic
e-mail: karel@zvara.cz

The Database of the Catalogue of Clinical Practice Guidelines Published via Internet in the Czech Language - The Current State

M. Zvolský¹

¹Centre of Biomedical Informatics, Institute of Computer Science AS CR, v.v.i.

Supervisor: Doc. Ing. Arnošt Veselý, Csc.

Summary

Since the nineties the publication activity of expert authorities in the field of Clinical practice guidelines (CPGs) has increased. During the first decade of the 21st century the Internet has become one of the most important publishing media for CPGs due to the popularity of sharing information in an electronic form. Nowadays, approximately 107 Czech expert societies publish about 731 CPGs on the Internet according to previous explorations of web sites of Czech expert medical societies and practical experience with previous operation of the Catalogue of Clinical Practice Guidelines. CPG documents can be used by physicians and other clinical employees in clinical praxis and education, by administrative authorities, by health policy makers or by patients. Information in these documents is also essential in the development of clinical information systems and decision support systems. The database of the Catalogue of Clinical Practice Guidelines takes a survey in the appearance of CPGs on free accessible web sites of expert societies or periodicals guaranteed by expert societies. The database stores bibliographical data, information about the recency and the quality of CPGs and information about the existence of web tools or formal models based on these CPGs.

The database contains 562 records of CPGs published by Czech expert medical societies. 212 records are accessible via the web interface of the Catalogue of Clinical Practice Guidelines.

Keywords: Internet, World Wide Web, database, clinical practice guideline, clinical practice, evidence-based medicine, formalisation, GLIF (Guideline Interchange Format), doctor of medicine, decision support systems

1. Introduction

1.1 Definition of Clinical practice guidelines in the Czech language

Clinical practice guidelines (CPGs) are an important part of the decision process on all levels of health care providing [1]. A recognized definition of these documents is they are "Systematically developed statements that can be used to accomplish appropriate specific health care decisions, services and outcomes" [2].

In spite of the big number of CPGs in Czech expert publications, they have often different designations such as "medical recommendation", "recommended procedure", "methodical instruction", "guideline", "standard" and others. The Czech Society of Cardiology uses both "guideline" (the word "guideline" is adopted into the Czech language) and "recommended practice" ("doporučený postup" in original) as synonyms. The Czech Medical Association of J. E. Purkyně uses terms "therapeutic standard" ("léčebný standard") and recommended practice ("doporučený postup"). The terms used started to be defined accurately in the Czech expert literature in recent years [1], [3], [4], [5].

1.2 Publication activity of expert medical societies

Documents defining health care in particular clinical relations are created on several levels. Documents with supreme legitimacy are published by the World Health Organization. Other documents published by international and national expert institutions have often geographically limited acceptance. Documents published by individual health care institutions have only a local sphere of action.

In the following text we use only CPGs with relevance on the territory of the Czech

Republic. They can be created by Czech national expert societies or translated or adopted from international CPG with wider acceptance.

Standards of health care are not prescribed by imperative legal rules in the Czech law [6]. Adequacy of the health care is evaluated (e.g. by a competent supervisory authority in case of a mistake) as „lege artis“ or „non lege artis“. Only an attending physician can make clinical decisions and selects the best clinical care. They decide not only on the basis of existing CPGs but they also judge the actual condition of the particular patient in its complexity.

Professional medical societies associated in the Czech Medical Association of Jan Evangelista Purkyně are reputable authorities for publishing CPGs in the Czech Republic. The present number of these societies is 107 [5]. Only some of them systematically develop documents of the normative nature defining administration of the health care in the given clinical conditions (see Table 1) [7]. The Czech Medical Chamber, Czech Dental Chamber, Ministry of Health, National Institute of Public Health or Chief Public Health Officer are other potential publishing authorities of CPGs in the Czech Republic.

There are several ways for CPG document publishing and dissemination. In a printed version CPGs can be an article in an expert periodical of the professional medical society (e.g. *Cor et Vasa*, Czech Society of Cardiology) or an individual brochure (e.g. series Recommended Diagnostic and Therapeutic Practice for General Practitioners, Society of General Practice).

Tab. 1. Numbers of CPGs electronically published via web sites of Czech professional medical societies or their official periodicals, year 2009 (showing only societies with eight or more documents).

Name of expert society	Number of CPGs
Society of General Practice	47
Czech Society of Cardiology	24
Czech Society of Dermatology and Venereology	36
Czech Society of Gastroenterology	14
Czech Society of Neurology	18
Czech Pneumological and Phthisiological Society	22
Czech Society of Angiology	8
Czech Diabetes Society	13
Czech Society of Hepatology	18
Czech Society for Oncology	79
Czech Rheumatological Society	9
Czech Society of Clinical Biochemistry	34
Czech Society for Atherosclerosis	10
Czech Medical Society for Infectious Diseases	19
Czech Society for Emergency and Disaster Medicine	17
Czech Society of Pathology	9
Czech Society for Transfusion Medicine	8
Czech Radiological Society	13
Czech Society of Hematology	14
Czech Society of Anaesthesiology and Intensive Care Medicine	13
Czech Medical Association of J.E. Purkyně	305

In an electronic version CPGs are articles in electronic versions of expert periodicals of professional medical societies (e.g. *Vnitřní lékařství*, The Czech Society of Internal Medicine) [8] or more often as electronic documents located on web sites of professional medical societies (e.g. Czech Society of Dermatology and Venereology) [9]. In all cases the electronic document in an electronic version of textual CPGs is most frequently presented in the Portable Document Format (PDF), the Hypertext Markup Language (HTML) format or the Microsoft Office Document (DOC) format. [7]

As professional societies develop CPGs independently of each other there is no

consensus about features of Czech CPGs and individual documents differ not only in formal aspects but also in the amount of information contained. CPGs of the Society of General Practice are an exception, because this society and an established Centre for Clinical Guidelines developed a basic methodology for clinical guidelines. They issue these CPGs in a compact edition of brochures accessible both in printed and electronic versions.

There is an effort for an introduction of standards of clinical care as a part of health care system of the Czech Republic. The National Reference Centre was founded for this purpose. A detailed methodology was assembled for the systematic

development of these standards. Individual standards are not open to the public at the present time [9].

1.3 Catalogue of Clinical Practice Guidelines - the history and goal of the project

The rich amount of published CPGs, their different forms, the amount of publishing institutions and the number of their web sites where these documents are stored in electronic versions cause problems in searching for the particular document for a specific clinical situation.

The problem of the quality of the looked-up documents arises when using full-text web search engines or general catalogues. Many different organisations and individuals such as health care providers, professional commercial portals, patients, special-interest associations or pharmaceutical companies can publish texts mentioning clinical situations, diagnostics and treatment. These documents may differ from the CPGs in effect or even contradict them.

Specialized portals dedicated to the issue of CPGs exist abroad. They often contain catalogues of electronically published CPGs. The best known CPGs catalogues are the National Guidelines Clearinghouse in the USA [10] and the National Library of Guidelines in the United Kingdom [11] both focused on CPGs in the English language. Catalogues focused on national CPGs arised also in other countries (e.g. <http://www.leitlinien.de/>, <http://www.nzgg.org.nz/>). The Guidelines International Network (G-I-N) connects national organisations into the international web. It also has a library with links to the on-line accessible CPGs [12]. Although Czech Národní referenční centrum is a member of G-I-N, there is no link to the Czech CPGs in the International Guideline Library or other foreign or international catalogues. Foreign web services in the field of CPGs have the disadvantage of foreign language web interface, which could be a barrier for some Czech users.

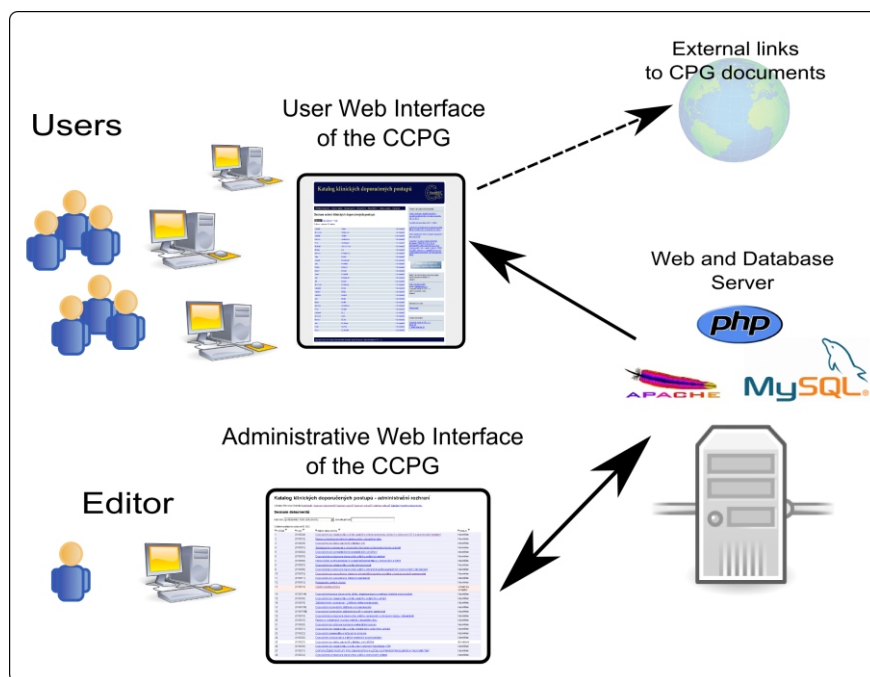


Fig. 1. Catalogue of Clinical Practice Guidelines.

a directory of these experts and institutions. In the future we assume to connect a number of links between each CPG document and external bibliographical links (e.g. into Medline or other bibliographical databases). This can create a useful source for educational purposes. Our database is ready for linking information about formal models of CPGs. These models are not included in today's Czech CPGs and they are not available from external sources yet except our previous efforts. Links between CPGs, classification and terminological systems together with future development of formal models of CPGs can compose a valuable informational source for decision support systems [14]. We also consider future extensions of the database with the extension of the Web 2.0 functions of the CCPG portal.

The development of the Catalogue of Clinical Practice Guidelines (CCPG) (former *Lékařské doporučené postupy v ČR* - Clinical Practice Guidelines in the Czech Republic) started in 2007 [13]. The absence of the general list of electronically published Czech CPGs was an impulse to start the development of our catalogue. We focused on CPGs in the Czech language or documents valid for the Czech Republic respectively, which are published via the Internet. We designed a web interface in the PHP scripting language, which allows to display records stored in the MySQL database and their listing according to various criteria. This freely accessible user web interface is placed on the following address <http://neo.euromise.cz/kkdp>.

Another web interface (administrative interface) is accessible after authentication and it allows adding and edition of records, communication with authors of CPGs and among editors, checking the validity of external links and other functions. The schematic structure of the CCPG project is described in Figure 1.

The database of gathered data about several CPG documents has a wider use. Because of a high amount of information about authors and professional societies, the database can serve as a source for

Tab. 2. Structure of the database table *_documents*.

Column name	Data type	Description
category_id	int(11)	category of the CPG document
document_abstract	Text	abstract of the CPG document
document_actualization	Date	date of the last actualization of the CPG document
document_authornote	Text	authors annotation of the CPG document
document_contact	Text	authors contact information
document_creation	Date	date of creation/authorization of the CPG document
document_id	int(11)	serial number of the CPG document
document_iid	varchar(20)	unique coding identification of the CPG document
document_inda	Datetime	date and time of the CPG record insertion
document_keywords	Text	key words for the CPG document
document_lcds	Datetime	date and time of the CPG record change
document_lcui	smallint(6)	identification of the user, who made the last change to the CPG record
document_note	Text	textual note to the CPG document
document_printed	Text	complete title of the printed version of the CPG document
document_state	int(11)	state of verification of the CPG document record – states whether the record is fully verified by the editor and should be displayed in the user interface
document_target_population	Text	target population for the CPG document
document_termination	Date	date of the CPG document withdrawal
document_title	Text	title of the CPG document
document_usid	smallint(6)	identification of the user, who created the CPG record
document_views	int(11)	number of views of the CPG document in the web interface of the CCPG
evidence_id	int(11)	lowest level of evidence used in the CPG document using the scale of the Oxford Centre for Evidence-Based Medicine
document_evidence	Text	definition and description of the levels of evidence used in the CPG document
recency_id	smallint(5)	recency type of the CPG document – concept of document, valid document or obsolete document

2. Methods

In the concept of the CCPG database we resulted partly from the statement of the Conference on Guideline Standardization (COGS) [15], partly from an inspiration with the National Guideline Clearinghouse [10] with comprehensive details about every CPG document in the „Guideline Summary“. These sources were confronted with the real situation of Czech CPG documents, which often do not contain basic identification specification such as the name of the author or the date of an issue. It was not possible to use developed tools for communication with CPG authors for completing CPG records. In the project of CCPG we presume using these tools for intensive communication and cooperation with CPG authors, but we do not dispose of resources comparable with the Agency for Healthcare Research and Quality in the United States, which operates the National Guideline Clearinghouse. It was our aim in the introductory phase to enter information, which is already contained in CPG documents or which can be easily found (e.g. related diagnoses in the ICD-10 classification). So we accentuated a structured identification specification about authors, time and expert destination of the CPG. We can define another information for example in the universal textual column `document_note` in the `_documents` table.

We have been developing the actual version 2.0 of the CCPG since the beginning of the year 2010. We have been using a MySQL version 5.0.41 database as a database environment. The database consists of 49 tables. 18 tables contain records about CPG documents, authors, authorities, used classification and thesaurus codes and auxiliary code-lists. 13 tables contain data used for web interface functions. 18 tables are binding tables.

Several codes of international classification and nomenclature systems can be assigned to the CPG document record in the CCPG database:

1. *Medical Subject Headings (MeSH) – complex vocabulary thesaurus of the National Library of Medicine*

One or more MeSH codes can be linked to each CPG document. MeSH code

describes the position of the term naming descriptor in the hierarchical structure. One descriptor can occur on different positions in the hierarchical structure - in this case the most matching hierarchical position is coded.

2. *The International Classification of Diseases and Related Health Problems 10th Revision (ICD-10)*

One or more three to four letter codes can be linked to each CPG document.

3. *Diagnosis-related group (DRG) - hospital cases classifying system*

One or more five-digit codes can be linked to each CPG document.

4. *SNOMED CT (Systematized Nomenclature of Medicine - Clinical Terms) - systematically organized collection of medical terminology*

One or more unique concepts of this system can be linked to each CPG document. SNOMED CT has not yet been translated into Czech and therefore no Czech CPG document defines related SNOMED CT concepts. We can find appropriate SNOMED CT concepts by translating key terms of the CPG into English or we can map above-mentioned classification into SNOMED CT.

Tab. 3. Structure of the database table `_linked_files`.

Column name	Data type	Description
<code>file_id</code>	<code>int(11)</code>	serial number of the record
<code>file_name</code>	<code>text</code>	link title
<code>file_link</code>	<code>text</code>	full definition of the hypertext link (URL)
<code>filetype_id</code>	<code>smallint(6)</code>	file type (podle externího číselníku)
<code>file_note</code>	<code>text</code>	link comment
<code>file_size</code>	<code>int(11)</code>	file size in kilobytes
<code>file_inda</code>	<code>datetime</code>	date and time of the link record insertion
<code>file_usid</code>	<code>int(11)</code>	identification of the user, who created the link record
<code>file_lcda</code>	<code>datetime</code>	date and time of the link record last change
<code>file_lcui</code>	<code>int(11)</code>	identification of the user, who last changed the link record
<code>file_status</code>	<code>smallint(6)</code>	recency type of the linked CPG document – concept of document, valid document or obsolete document
<code>file_show</code>	<code>boolean</code>	marks displaying of the link in the user interface
<code>file_actualization</code>	<code>date</code>	date of the last actualization of the linked document
<code>broken_link</code>	<code>boolean</code>	mark of nonfunctional link made by an automated evaluation tool

The editor assigns codes of classification systems to the CPG document in the administrative interface of the CCPG. It is possible to assign multiple codes to each document. They can be found as a part of the document in a standalone list in some cases, or it is possible to find them actively disseminated in the whole text. The document does not contain codes of classification systems directly in most cases. Than it is necessary to extract key terms from the document and find appropriate codes manually or through external tools.

The Catalogue of Clinical Practice Guidelines uses also other code lists:

- The list of authorities (primarily professional medical societies), which is continuously extended when adding new CPG documents.
- The list of CPG authors is also continuously extended when adding new CPG documents.
- The list of medical and healthcare specialties.
- Code-lists for geographical competency, language versions or file types.

The main database table `_documents` consists of 23 columns defining the CPG document (see Table 2). Further information is provided by binding tables connecting above-mentioned code-lists and the table of external links.

Fig. 2. Screenshot generated by a web browser from the URL http://neo.euromise.cz/kkdp/?akce=mkn_strom shows browsing through hierarchical structure of the ICD classification with hypertext links to ICD codes definitions on the web page of the Institute of Health Information and Statistics of the Czech Republic.



browsing web sites of professional medical societies (using the list placed on the web page of the Czech Medical Association of J. E. Purkyně). We insert the structured description of all CPG documents found on these sites and meeting the above criteria into the CCPG database using the administrative interface. Another way of enlisting a CPG document is using the announcement message generated from the web form, which is a part of the CCPG user web interface. If the visitors of CCPG know the address of an unlisted document, they can announce it to editors using this form.

After creating a new record of the CPG, the editor searches for the maximum amount of information available about the CPG document and adds it to the database using the administrative interface. In the future we count on an active cooperation with CPG document authors, who can complete or confirm the data gathered. Every record has a variable describing the phase of the processing.

This table named `_linked_files` defines addresses of CPG documents as the Uniform Resource Identifiers. This table consists of other 14 columns defining of each link in detail (see Table 3).

Fig. 3. Screenshot generated by a web browser from the URL <http://neo.euromise.cz/kkdp/?akce=detail&dokument=46> - CPG document record details with link to the full text document URL.



Following criteria were stated for the enlisting the CPG document into the CCPG database:

- The CPG document is accessible in the electronic format via Internet.
- The CPG document is granted by at least one professional authority. It is linked from the web site of the granting authority or the official periodical of this authority. It is placed on the same web site.
- The CPG document is defined in its title or related description as clinical practice guideline, guideline, recommendation, recommended practice, standard, methodical direction or protocol.
- The CPG document is accessible free of charge in the public part of the web presentation with no registration required. It is not password protected and it has no other protection.

Two possible ways of selecting the documents suitable for enlisting in CCPG take place. Since 2007 we have been

In compliance with this variable the record is listed and displayed on the web interface indicating the completeness and the recency of the record.

We developed a tool for automated browsing and evaluation of recorded links to the CPG documents. This tool analyses returned HTTP status codes for each link. The link is marked as an error and it is highlighted for the editor to correct it, when the error HTTP status code is returned. The editor then analyses an error and they correct or delete the link in the database.

The user interface of the CCPG offers the list of CPG document records ordered by the title or by the date of the last modification. Other possible lists are orders by authors (alphabetically) or by authorities (professional medical societies). Users can browse through the hierarchical tree of MeSH or ICD when searching for the target document by topic (see Figure 2). It is possible to choose and display the detailed information about the particular CPG document, where all available data about the document are printed including hypertext links to the full text document URL (see Figure 3).

3. Results

On September 1st, 2010 the database of the Catalogue of Clinical Practice Guidelines contained 562 records of Czech Clinical Practice Guidelines. 12 records were marked as deleted/not existing. 212 records were marked as verified and they were displayed in the user web interface of the CCPG.

The database contained 640 records of linked files of CPG documents. 17 of them were marked as an incorrect link.

On the same date the database contained 491 records of authors of the CPG documents and 91 records of publishing professional medical societies.

Only records of CPGs marked as verified and authors and authorities related to these records are displayed in the freely accessible user web interface of the CCPG. Therefore this interface of the catalogue lists only 212 records of CPG, 430 records of authors and 46 records of authorities.

4. Discussion

The database of the CCPG in the current state maps the majority of current publication activities in the field of CPGs accessible via the Internet. The total number of these documents is a minimum of 731 of linked files [7]. 640 of them are recorded in our database (or more precisely 623 existing files).

Our catalogue contains fewer records in comparison with some foreign catalogues of clinical practice guidelines as the Guidelines International Network, the National Guideline Clearinghouse or the National Library of Guidelines [12], [10], [11]. This difference is caused mainly by focusing on documents in the Czech language. The amount of CPG documents in English is much bigger. Not all records in above mentioned catalogues include the reference to the on-line full text version of the CPG document. Comparison with foreign catalogues focused only on the national set of CPG documents or limited on one development organization (e. g. the New Zealand Guidelines Group [16] or the Scottish Intercollegiate Guidelines Network [17]) is more advantageous in the number of records (see Table 3). Our catalogue has more possibilities in ordering records by different categories contrary to foreign catalogues.

Many Czech CPG documents are so incomplete or they do not contain sufficient identification specification that it is necessary to search hard for the missing specification data.

Only 212 records are marked as verified due to limited resources assigned to the maintenance of the catalogue.

Czech CPG documents in most cases do not contain codes of classification and nomenclature systems, but they contain textual information about diagnoses, etiological agents or anatomical structures, which can be found in these classifications. Manual browsing in the classification and searching for the code by the editor is less accurate than these being specified by the author of the CPG. The inaccuracy can arise when the term contained in the textual CPG does not correspond precisely with the term from the classification or it has multiple meanings in the hierarchical structure.

The database of the CCPG offers links to two classifications and a thesaurus commonly used in the Czech health care system (MeSH, ICD, DRG). SNOMED CT has not a Czech translation yet and is of a limited use, but it is the most complex nomenclature system. Filling codes of all four systems in cope with several problems. We have often insufficient information in the source text of the CPG. Terms in these four systems are not always fully matching. When filling codes of these systems properly in all records of CPG the potential for connecting to health information systems (ICD, DRG, SNOMED CT) and bibliographical systems would increase.

Tab. 4. - Comparison of some on-line catalogues of CPG.

Catalogue	Number of records	Records ordered by
Catalogue of Clinical Practice Guidelines	623 (212 validated)	author, expert society, MeSH, ICD
Guidelines International Network	7100	MeSH
National Guideline Clearinghouse	2549	MeSH, developer
National Library of Guidelines	3220	specialization
New Zealand Guidelines Group	74	therapeutic category, developer
Scottish Intercollegiate Guidelines Network	120	topic

The database should also contain information about formalised models of CPGs in the form of links to descriptions of these models or to web applications using or displaying these models.

Nowadays, the database does not contain this information because formal models of CPGs do not exist or they are not freely accessible via Internet. In the future the CCPG could serve as an on-line repository for the information about these models.

The database of the CCPG is designed for storing the information about CPG documents published in the Czech language. CPG documents with a wider geographical effect can be valid for the Czech Republic (e.g. documents issued by the World Health Organisation or international professional societies). These documents can also have foreign language versions. The future development of the CCPG will respect and support various language versions of the CPG documents.

The feature of adding information about relevant literature with linking to bibliographical sources like the Medline is another planned modification of the CCPG. The database of the CCPG will be adjusted also in other aspects according to the actual needs.

The aim of the CCPG project is to store information about CPG documents and completely cover all publication activities of professional authorities in the field of CPGs. For this purpose we assume to develop tools for automating processes of checking the link validity and newly also browsing web sites for new CPG documents. We will use tools for better communication with CPG authors. We will try to establish a professional community around the project of the CCPG.

5. Conclusion

Since 2007 we have developed the Catalogue of Clinical Practice Guidelines for the purpose of collecting information about Clinical practice guideline documents available for the Czech Republic. The component of this project is a database of CPG records in the MySQL format, which can have further usage in decision support systems. On September 1st, 2010 the database included 550 records of active CPG documents in the total of 623 valid files. The database will be extended and adjusted to the publication activity of professional authorities in the Czech Republic.

Acknowledgment

*CCPG development is supported by the project 1M06014 of the Ministry of Education, Youth and Sports CR.
This work was supported by the grant SVV-2010- 265 513.*

References

- [1] Ličeník R.: Klinické doporučené postupy. Výukový portál LF UP Olomouc [online] 20.11.2009, poslední aktualizace 20.11.2009 [cit. 2010-09-24], dostupný z: <http://mefanet.upol.cz/clanky.php?aid=38>
- [2] Field M.J., Lohr K.N.: Guidelines for Clinical Practice: From Development to Use, NATIONAL ACADEMY PRESS, 1992
- [3] Suchý M.: Příručka pro autory a oponenty vývoje klinických doporučených postupů, Národní referenční centrum [online] 1. 9. 2010, dostupný z: http://www.sopr.cz/standardy/dokumenty/dokument_open.php?doc=3
- [4] Česká kardiologická společnost, <http://www.kardio.cz>
- [5] Česká lékařská společnost Jana Evangelisty Purkyně, <http://www.cls.cz>
- [6] Těšínová J.: Právní aspekty poskytování zdravotní péče v ČR, Medical Tribune 18/2010, dostupný z: <http://www.tribune.cz/clanek/18816-pravni-aspekty-poskytovani-zdravotni-pece-v-cr>

- [7] Zvolský M.: Základní parametry dokumentů doporučených postupů českých lékařských společností publikovaných prostřednictvím Internetu, Doktorandské dny'09, MATFYZPRESS; 2009, dostupný z: <http://www.cs.cas.cz/hakl/doktorandsky-den/files/2009/sbornik-dd-2009.pdf>
- [8] Vnitřní lékařství, <http://www.vnitrmilekarstvi.cz>
- [9] Národní referenční centrum, <http://www.nrc.cz>
- [10] National Guideline Clearinghouse, <http://www.guidelines.gov>
- [11] National Library of Guidelines, <http://www.library.nhs.uk/guidelinesfinder>
- [12] Guidelines International Network, <http://www.g-i-n.net>
- [13] Zvolský M.: Katalog lékařských doporučených postupů v ČR, Doktorandské dny'08, MATFYZPRESS; 2008, dostupný z: <http://www.cs.cas.cz/hakl/doktorandsky-den/files/2008/sbornik-dd-2008.pdf>
- [14] Buchtela D., Peleska J., Zvolský M., Zvarova J.: Medical knowledge representation system, Stud Health Technol Inform. 2008;136:377-82.
- [15] Shiffman R.N., Shekelle P., Overhage J.M., Slutsky J., Grimshaw J., Deshpande A.M. Standardized reporting of clinical practice guidelines: a proposal from the Conference on Guideline Standardization. Ann Intern Med. 2003 Sep 16;139(6):493-8
- [16] New Zealand Guidelines Group, <http://www.nzgg.org.nz>
- [17] Scottish Intercollegiate Guidelines Network, <http://www.sign.ac.uk>

Contact

MUDr. Miroslav Zvolský

Centre of Biomedical Informatics
Institute of Computer Science AS CR
Pod Vodárenskou věží 2
182 07 Prague 8
Czech Republic
e-mail: zvolsky@euromise.cz



EJBI 2010

ISSN 1801-5603

European Journal for Biomedical Informatics

Volume 6 (2010), Issue 1

Czech version

Cooperating journal – Lékař a technika

Editor-in-Chief: Karel Roubík (Czech Republic)

Publisher: Společnost biomedicínského inženýrství a lékařské informatiky ČLS JEP
a Fakulta biomedicínského inženýrství ČVUT v Praze

An Official Journal of the European Federation of Medical Informatics

www.ejbi.eu



Content

Czech version

- cs 2 [Předmluva](#)
Svačina Š., Zvárová J.
- cs 3 - 7 [Problematika mozkových nádorů astroglální řady](#)
Bielniková H., Buzrla P., Dvořáčková J.
- cs 8 - 14 [Multiagentní architektura založená na procesech aplikovaná na doménu domácí péče](#)
Bošanský B.
- cs 15 - 19 [3D-CT rekonstrukce náhrady temporomandibulárního kloubu před a po terapii](#)
Feltová E., Dostálová T., Kašparová M., Daněk J., Hlíňáková P., Hubáček M., Nedoma J.
- cs 20 - 26 [Metody analýzy přežití](#)
Fürstová J.
- cs 27 - 32 [Elektronická zdravotnická dokumentace pro onemocnění temporomandibulárního kloubu – Podpora v rozhodování o terapeutickém procesu](#)
Hippmann R., Nagy M., Dostálová T., Zvárová J., Seydlová M., Feltlová E.
- cs 33 - 37 [Autoimunitní thyroditida – porovnání vyšetřujícího lékaře a automatické klasifikace, vyhodnocení variability lékařů](#)
Holinka Š., Smutek D.
- cs 38 - 42 [Vztah zobecněné entropie a Bayesovské pravděpodobnosti chyby](#)
Horáček M.
- cs 43 - 52 [Metody pro tvorbu časově závislých referenčních intervalů a kvantilových křivek pro hodnocení velikosti plodu](#)
Hynek M.
- cs 53 - 57 [Infrastruktura pro ukládání dat a výpočty v biomedicinském výzkumu](#)
Kulhánek T.
- cs 58 - 65 [Jazyk českých lékařských zpráv a klasifikační systémy v medicíně](#)
Přečková P.
- cs 66 - 72 [Porovnávání internetových stránek a lékařských doporučených postupů s využitím databází řízených medicínských slovníků v kontextu hodnocení kvality webových zdrojů](#)
Rak D.
- cs 73 - 75 [Zrychlení algoritmu pro nalezení optimální šířky jádra v analýze sledu akčních potenciálů](#)
Šanda P.
- cs 76 - 77 [Jednoduchá ekonomická analýza diagnostiky nemocí štítné žlázy při screeningu těhotných žen](#)
Telička Z., Jiskra J., Springer D., Límanová Z.
- cs 78 - 82 [Identifikace jednotek a dalších termínů v českých lékařských zprávách](#)
Zvára K., Kašpar V.
- cs 83 - 89 [Stav databáze Katalogu klinických doporučených postupů publikovaných prostřednictvím internetu v českém jazyce](#)
Zvolský M.

Předmluva

Od moderních informačních systémů se očekává, že dovedou efektivně pracovat s informacemi a vzájemně si je vyměňovat. Pokud jsou informace sdíleny pomocí volného textu, je základním předpokladem pro sémantickou interoperabilitu přístup k jejich významu. Sémantická interoperabilita a schopnost pomocí technických prostředků získat konkrétní informaci jsou základní podmínky pro využití dalších technologií telemedicíny a eZdraví. Schopnost systémů navzájem porozumět předávaným datům (sémantická interoperabilita) vyžaduje používání stejného pojmosloví (tedy klasifikačních systémů a nomenklatur) a využívání stejného jazyka pro komunikaci a její zápis (datové standardy). Využívání stejného pojmosloví může nejlépe zajistit používání existujících norem vytvořených pro náš region. Existující normy (např. EN 13606) předpokládají využití jednoznačných

a celosvětově unikátně definovaných termínů, které lze bez větších obtíží obecně převést do jiných klasifikací (např. prostřednictvím systému Unified Medical Language System). Zřejmě nejlépe použitelným obecným klasifikačním systémem pro zdravotnictví je systém SNOMED CT. Ten vznikl spojením amerického systému SNOMED (vytvořeném Asociací amerických patologů) a britského systému Clinical Terms („Read Codes“). V souvislosti s tímto spojením došlo ke vzniku sdružení International Health Terminology Standards Development Organisation (IHTSDO) se sídlem v Dánsku.

V rámci projektu specifického výzkumu Sémantická interoperabilita v biomedicině a zdravotnictví studenti doktorského studia na 1. lékařské fakultě Univerzity Karlovy v Praze vypracovali jako hlavní autoři

odbornou práci v angličtině a v češtině k tématům jejich disertačních prací. Práce uváděné v tomto sborníku jsou české verze prací, které prošly recenzním řízením a budou do konce roku 2010 publikovány jak v české tak i v anglické verzi v internetovém časopisu European Journal for Biomedical Informatics (www.ejbi.eu). Uvedený internetový časopis je zařazen do seznamu recenzovaných neimpaktovaných časopisů vydávaných v České republice, který schválila Rada vlády. Přílohou sborníku semináře je i aktuální verze výkladového anglicko-českého slovníku sestaveného z podkladů doktorandů a dalších informačních zdrojů.

Prof. MUDr. Štěpán Svačina, DrSc., MBA
Prof. RNDr. Jana Zvárová, DrSc.
editoři

Problematika mozkových nádorů astroglální řady

H. Bielníková¹, P. Buzrla¹, J. Dvořáčková¹

¹Ústav patologie, Fakultní nemocnice Ostrava

Školitel: Doc. RNDr. Ivan Mazura, CSc.

Souhrn

Astroglální tumory jsou skupinou nádorů derivujících se z buněk astroglální řady. Jsou to jedny z nejčastěji diagnostikovaných tumorů nervového systému. Tato skupina nádorů je v mnoha aspektech různorodá. Nádory se od sebe liší nejen morfologicky, ale také růstovým potenciálem, rozsahem invazivního růstu, svou lokalizací v CNS i klinickým průběhem v důsledku jejich progresivních vlastností. Různorodost těchto novotvarů je dána sekvencí a typem genetických alterací v průběhu procesu nádorové transformace. Podle současné užívané klasifikace můžeme astroglální nádory rozdělit na benigní formy I.-II. (low grade glioma-I. pilocytární astrocytom, II. difúzní astrocytom) a maligní formy III.-IV. (high grade glioma-III. anaplastický astrocytom, IV. glioblastom). Pro určení, zda-li je nádor spíše benigní či maligní povahy je z histologického hlediska rozhodující množství vazoproliferativních a nekrotických změn spolu s počtem buněčných dělení. Z hlediska molekulárně-genetického jsou důležité typy a počty kumulujících se genetických alterací. Jedny z podstatných alterací nastávají v genech p53, PTEN a EGFR.

Klíčová slova: astrocytom, glioblastom, imunohistochemie, molekulárně-biologické metody, mutace, tumorsupresorové geny, amplifikace, gen p53, gen PTEN, amplifikace EGFR faktoru

1. Úvod

Astroglální tumory jsou jedny z nejčastějších primárních mozkových tumorů intrakraniální oblasti tvořící, podle údajů v literatuře, více jak 60 % případů primárních nádorů mozku. Jejich klasifikace je nejčastěji hodnocena na základě mezinárodně uznávaných kritérií stanovených Světovou zdravotnickou organizací (WHO). Vzhledem k různorodosti histologického obrazu a odlišnému biologickému chování se tyto nádory dělí do několika typů. Lokalizace astroglálních tumorů může být kdekoliv v CNS,

nejčastější výskyt však bývá v mozkových hemisférách. Růstová povaha těchto nádorů je převážně infiltrativní a jejich nebezpečím je tendence přechodu do malignějších forem. Tato skupina nádorů je ve větší míře diagnostikovaná u dospělých, u některých typů byl také prokázán větší výskyt u mužů [1], [2].

Stupně malignity a diferenciací nádoru určujeme podle gradingu, který je významným indikátorem prognózy a je důležitý pro stanovení léčebných postupů. Mezi kritéria důležitá pro určení gradingu patří proliferativní aktivita buněk a stupeň buněčné diferenciací (v histopatologickém obraze je sledována zralost nádorové tkáně – stupeň diferenciací), přítomnost buněčných a jaderných atypií, mitotická aktivita, angiogeneze a přítomnost nekrotických změn [1], [2]. Podle biologické povahy, můžeme astroglální tumory rozdělit na relativně benigní a maligní formy (Tab. 1.).

Tab. 1. Grading astroglálních tumorů.

Grade	Název mozkového nádoru	Povaha nádoru
I	Pilocytární astrocytom	relativně benigní
II	Difúzní (Low-grade) astrocytom	relativně benigní (s tendencí maligní progresí)
III	Anaplastický astrocytom	maligní
IV	Glioblastom	vysoce maligní

Z histopatologického hlediska je pro určení povahy tumoru rozhodující množství vazoproliferativních a nekrotických změn, míra diferenciací buněk, počet buněčných dělení a jaderné a buněčné atypie. Z hlediska molekulárně-genetického je důležitá přítomnost charakteristických genetických alterací a jejich množství.

Jedny z podstatných změn nastávají v genech p53, PTEN, EGFR, Rb a p16.

1.1 Současná klasifikace astroglálních nádorů

Pilocytární astrocytom je podle WHO klasifikace Grade I tumor. Tento typ nádoru se vyskytuje převážně u dětí a mladých dospělých v různých lokalizacích CNS, často však postihuje 5. oblast mozečku a optického nervu. Jedná se o ohraničený, pomalu rostoucí tumor, spíše benigní povahy, prognosticky příznivý. V histologickém obraze lze pozorovat bifazické uspořádání tumoru s oblastmi hustě nakupených GFAP (gliální fibrilární kyselý protein) pozitivních bipolárních astrocytů a eosinofilních Rosenthalových vláken. Dále jsou zde přítomny oblasti řídké celulózy, pěníte až mikrocystické s GFAP slabě pozitivními protoplasmatickými astrocyty bez fibril. Častá je přítomnost regresivních změn v podobě hyalinizovaných cév a kalcifikací. V menší míře se objevují jaderné pleomorfie, které však nejsou prognosticky významné. Nekrotické změny, angiogeneze či mitózy u pilocytárního tumoru pozorovány nebývají [1], [2].

Druhým tumorem podle gradingu je difúzní nebo také low-grade astrocytom. Tento tumor bývá dobře diferencovaný, pomalu rostoucí, má však povahu infiltrativního růstu s tendencí maligní progresí. Postihuje převážně dospělé mladšího věku. Mitotická aktivita difúzního astrocytomu je oproti vyšším gradům nízká. Morfologicky jsou zřetelné nádorové populace tvořené astrocytárními buňkami, obtížně odlišitelnými od buněk nenádorových. I přesto lze na základě histologického obrazu rozlišit tři typy low-grade astrocytomů [1], [2].

Nejčastějším je fibrilární astrocytom, kdy v buňkách s nízkou mitotickou aktivitou nacházíme jaderné atypie.

Typickým znakem bývá tvorba mikrocyst s mukoidním obsahem a přítomnost lymfocytárního infiltrátu v blízkosti cév. Při imunohistochemickém průkazu nádorové buňky exprimují GFAP (gliální fibrilární kyselý protein), S100 a vimentin. Proliferační marker Ki67 vykazuje nízkou expresi.

Druhým typem je gemistocytární astrocytom. Nádorové buňky bývají poměrně větší s hojnou, eosinofilní cytoplasmou a přítomností perivazálního lymfatického infiltrátu. Jádra s drobnými jádérky jsou uložena excentricky. Imunofenotyp i proliferace je shodná s předchozím typem.

Posledním typem low-grade astrocytomu, podle histologického obrazu, je méně častý protoplasmatický astrocytom. Populace nádorových buněk je řídká, buňky jsou drobného tvaru s oválnými jádry a nízkým obsahem gliálních filament.

Tvorba mikrocyst a mukoidní degenerace jsou charakteristickým znakem tohoto astrocytomu, imunohistochemicky lze stanovit fokálně GFAP.

Low-grade astrocytomy jsou z prognostického hlediska příznivé, průměrná doba přežití se pohybuje mezi šesti až osmi lety. Jedním z prognostických znaků je proliferace Ki67, kdy při expresi více než 5 % populace nádorových buněk je prognóza nádoru horší. Přítomnost perivaskulárního lymfoidního infiltrátu a mikrocyst je známkou příznivější prognózy [1], [2].

Anaplastický astrocytom vykazuje již vysokou proliferaci a jeho biologická povaha je maligní. Nádor se vyskytuje u dospělých, ve větší míře u mužské populace, mezi čtvrtou a pátou dekadou života. Grading nádoru je III. Tyto nádory jsou výrazně buněčnější než předchozí typy, jádra buněk jsou nestejně velká (anizonukleóza), jádérka jsou velmi výrazná. Mitotická aktivita je zvýšená s přítomností atypických mitóz. U tohoto typu začíná být patrná angiogeneze (novotvorba cév), oproti glioblastomu však nebývají přítomny nekrózy. Imunohistochemicky lze stanovit expresi S100 proteinu a vimentinu, exprese GFAP je kolísavá. Exprese proliferace markeru

Ki67 je obvykle mezi 5-10 %. Prognóza u tohoto typu nádoru je špatná, kdy průměrná doba přežití se pohybuje mezi dvěma až pěti lety. Tumor může vycházet z původně diagnostikovaného difúzního astrocytomu nebo může vzniknout *de novo*. Buňky ztrácejí diferenciační schopnosti a mají tendenci směřovat k nediferencovaným stádiím a neregulované proliferaci. Výsledkem většinou bývá přechod do malignější formy, glioblastomu [1], [2], [3].

Glioblastom je nejvíce frekventovaným zhoubným, mozkovým tumorem u dospělých ve věku mezi 45 až 75 lety, s převažujícím výskytem u mužů [2], [4]. Z hlediska vzniku se rozlišují dva typy nádorů. Primární glioblastom bývá častější, vzniká *de novo*, bez výskytu předchozích nádorových lézí. Nástup i klinický průběh onemocnění je rychlý. Druhý typ, sekundární glioblastom, vzniká pozvolně, maligní progresí z anaplastického astrocytomu či jiného gliálního tumoru. Histologicky i topograficky jsou glioblastomy velmi variabilní. Glioblastom je astrocytom grade IV, buňky vykazují nízký stupeň zralosti s projevy anaplazie, nádor je vysoce buněčný, tvořený nediferencovanými pleomorfními elementy. Přítomny jsou také obrovské buňky s velkými jádry a atypické mitózy. Typicky jsou přítomna ložiska (lokální i masivní) nekrózy obzvláště v centru nádoru, kdy nádorové buňky tvoří typické palisádové struktury na periferii nekrózy. Novotvorba krevních cév je rozsáhlá a hojná. Imunohistochemickými markery jsou vimentin a CD99, exprese GFAP je variabilní. Proliferační marker Ki67 vykazuje pozitivní expresi v přibližně 15 až 20 % [1], [2], [3], [4].

Průměrná doba přežití pacientů u nichž byl glioblastom diagnostikován je 1 rok od diagnózy. Z prognostického hlediska je významný věk pacienta, proliferace, aktivita nádoru a míra exprese EGFR. Čím je vyšší věk, proliferace i exprese EGFR, tím je prognóza horší [1], [2], [3], [4].

1.2 Genetické alterace nádorů astroglální řady

Díky molekulárně-biologickým metodám byla u astroglálních tumorů identifikována řada genetických alterací, které souvisejí

se vznikem a rozvojem těchto nádorů. Propuknutí nádoru většinou předchází nějaké iniciační poškození genu a následně se dále kumulují defekty, které vedou k progresi nádoru.

U astrocytárních tumorů je většinou tímto iniciačním bodem mutace genu p53. Tento gen je lokalizován na krátkém raménku chromozomu 17 (17p13.). Defekty tohoto genu patří v lidských nádorech mezi jedny z nejčastějších. Ve většině případů se jedná o vzniklé mutace obou alel tohoto genu, byly však popsány i případy zděděné mutace jedné alely. Produktem genu p53 je transkripční faktor, který za normálních podmínek stimuluje expresi dalších genů. Je důležitým tumorsupresorovým genem, který se podílí na regulaci buněčného cyklu a hraje důležitou roli při buněčném stresu. Blokuje buněčný cyklus a dává buňce čas na vyhodnocení stupně poškození a případnou následnou opravu DNA. Jsou-li vzniklé defekty neopravitelné spouští p53 mechanismy apoptózy. Defekty tohoto genu spolu s nadměrnou expresí PDGF (Platelet Derived Growth Factor) jsou spojovány se vznikem difúzních astrocytomů, které mohou dále progradovat do anaplastického astrocytomu až sekundárního glioblastomu [2], [3], [6].

Progrese astrocytomů směrem k vyššímu stupni malignity je dále zapříčiněna ztrátou genu Rb. Retinoblastomový gen (Rb) se nachází na dlouhém raménku chromozomu 13 (13q14). Produktem genu je jaderný fosfoprotein, který za normálních podmínek kontroluje transkripci jiných genů podílejících se na regulaci buněčného dělení. Stejně jako p53 má tumorsupresorovou funkci [2], [3].

Pilocytární astrocytom, který má Grade I a relativně benigní povahu se vyvíjí samostatnou genetickou dráhou po ztrátě genu NF1. Lokalizace NF1 je na dlouhém raménku chromozomu 17 (17q11). Produktem genu je protein - neurofibromin, který je negativním regulátorem proteinu p21^{ras} (má významnou roli v buněčné komunikaci a signalizaci) [2], [3], [6].

V případě primárního a sekundárního glioblastomu jsou přítomné defekty genů odlišné.

Primární glioblastom vznikající *de novo* má separátní vývojovou dráhu, kdy dochází k přímé transformaci ve vysoce maligní fenotyp bez mezistádií. Specifickými defekty primárního glioblastomu jsou delece genu p16, mutace genu PTEN, ztráta heterozygotnosti na chromozomu 10, mutace a amplifikace genu pro receptor epidermálního růstového faktoru (EGFR) a amplifikace MDM2 [1], [2], [3]. Gen p16 kóduje inhibitor cyklin dependentních kináz 4 a 6 a nachází se na krátkém raménku chromozomu 9 (9p21). Supresorový gen PTEN je lokalizován na dlouhém raménku chromozomu 10 (10q23.3). Jeho funkcí je blokáda buněčného cyklu v G₁ fázi. Právě poruchy tohoto genu jsou vázány s progresí astrocytomů nižších stádií k malignějším formám. Přítomnost této mutace je spojována s horší prognózou nádoru [2], [3], [6], [7], [8], [9], [10]. Gen pro EGFR receptor leží na krátkém raménku chromozomu 7 (7p12-14) a je nejčastěji amplifikovaným onkogenem u primárních glioblastomů. Jeho amplifikace pravděpodobně hraje úlohu v difúzní infiltraci glioblastomů do okolní tkáně. Vysoké počty amplifikací jsou spojovány s horší prognózou tumoru [2], [3], [4], [5], [6], [11].

Další amplifikovaný gen MDM2 je lokalizován na dlouhém raménku chromozomu 12 (12q13-14). Tento gen kóduje protein, jehož funkcí je negativní regulace apoptózy přes proteiny p53 a Rb [6].

V případě sekundárního glioblastomu, který se vyvíjí cestou přes difúzní astrocytom a anaplastický astrocytom, jsou nacházené mutace totožné s mutacemi u těchto nádorů nižších stupňů. V menší míře se mohou objevovat amplifikace genu pro EGFR receptor, delece genu p 16 či mutace v genu PTEN. Stejně, jako u primárního glioblastomu, je však ve velké míře přítomna ztráta heterozygotnosti na chromozomu 10 [2], [3], [6].

2. Současné přístupy v diagnostice

Nádorová tkáň mozku bývá pro bioptické vyšetření získávána klinikem při odběru pomocí kanyly. Získaný materiál může být k histologickému vyšetření dodán v nativní formě (odeslání **patologovi** ihned po odběru, uložení ve vlhké komůrce –

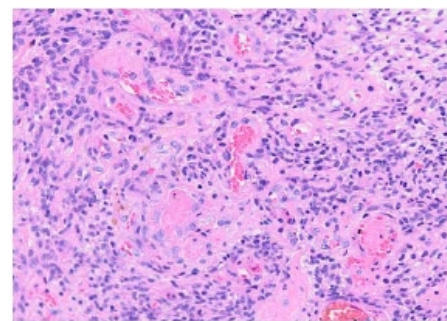
uzavřená nádobka se čtverečkem gázy namočené ve fyziologickém roztoku) nebo ve fixačním roztoku. Důležitým aspektem pro diagnostiku je správnost odběru, kdy při špatné lokalizaci může dojít k odebrání nenádorové tkáně. Důsledkem toho jsou zbytečně provedená nákladná a časově náročná vyšetření. Kritickým krokem je také zpracování materiálu, kdy je snahou předejít znehodnocení získané tkáně a zabránění vzniku artefaktů.

Molekulárně-biologickými metodami lze vyšetřit „čerstvou“ tkáň získanou při bioptickém odběru, stejně jako tkáň, která již byla zpracována metodou parafinových bloků.

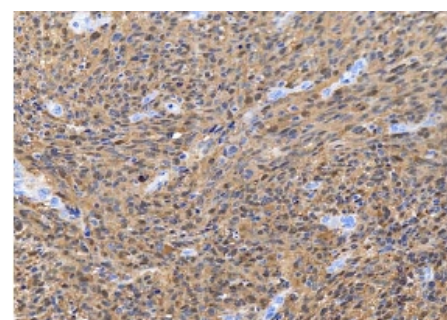
Hlavní diagnostické metody, pro určení astroglálních tumorů, jsou histologické a imunohistochemické vyšetření. Histologické vyšetření poskytne morfologický obraz nádoru (Obr. 1), imunohistochemické vyšetření je doplňující, avšak nedílnou součástí bioptické diagnostiky. V problematice astroglálních tumorů jsou nejvíce používanými markery vimentin, S100 protein (Obr. 2), gliální fibrilární kyselý protein- GFAP (Obr. 3) a proliferační marker Ki67 (Obr. 4). Markery pojivové tkáně, vimentin a S100 protein, jsou proteiny intermediárních filament. Vimentin se během vývoje exprimuje v buňkách a tkáních embryí, u dospělých organismů v mezenchymálních buňkách. V nezralých gliových elementech je předchůdcem gliálního fibrilárního kyselého proteinu (GFAP) a rovněž je spolu s GFAP koexprimován ve zralých astrocytech i jejich nádorových derivátech. S100 protein je kyselý protein centrálního i periferního nervového systému, avšak zcela specifickým proteinem nervové tkáně není, jelikož jeho přítomnost byla prokázána i v jiných tkáních [12].

Ki67 je markrem rychlosti buněčné proliferace. Reaguje s lidským jaderným antigenem exprimovaným v proliferujících buňkách všech fází buněčného cyklu, kromě fáze G₀ [12].

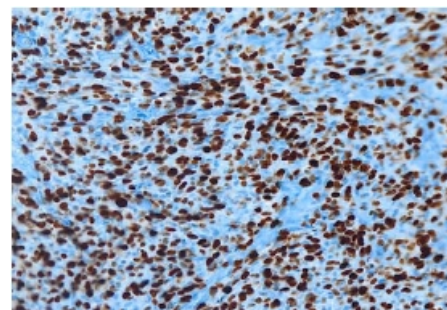
Imunohistochemické metody tak umožňují provést základní typizaci tkání a jejich derivátů.



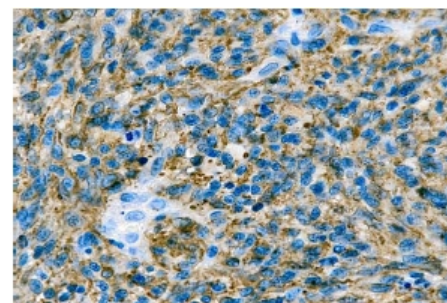
Obr. 1. Histologický obraz glioblastomu (HE, 400x, Ústav patologie, Fakultní nemocnice Ostrava).



Obr. 2. Expresa markeru S100 v tkáni glioblastomu (S100, 200x, Ústav patologie, Fakultní nemocnice Ostrava).



Obr. 3. Expresa proliferačního markeru Ki67 v tkáni glioblastomu (Ki67, 200x, Ústav patologie).



Obr. 4. Expresa GFAP proteinu v tkáni glioblastomu (GFAP, 400x, Ústav patologie, Fakultní nemocnice Ostrava).

S objevem a rozvojem nových metod, hlavně molekulárně-biologických, se posunuly možnosti diagnostiky dále. Tyto nové metody jsou dalším doplňkem v diagnostice a umožňují nám získat informace, které dané onemocnění více specifikují a upřesní. Lze jimi analyzovat známé mutace a alterace v genech, o nichž je známo, že jejich poškození může souviset se vznikem a rozvojem nádoru. Vyšetření těchto mutací se stává důležitým prognostickým faktorem, který je podstatný pro odhad dalšího průběhu onemocnění a pro případné zahájení adekvátní léčby. Stejně tak nám umožňují objevovat dosud nové, neznámé mutace, alterace a další faktory, které se na rozvoji onemocnění mohou podílet. Mezi často používané molekulárně-biologické metody v patologii patří hybridizace, elektroforéza, amplifikační metody (PCR – polymerázová řetězová reakce), sekvenování, micro-array analýzy apod.

3. Možnosti evidence a zpracování získaných dat

V průběhu každého výzkumu či studie dochází k produkci ohromného množství informací, kdy je nutno tyto získané informace nějakým způsobem zpracovat. V případě astroglálních tumorů můžeme získat informace týkající se průběhu onemocnění, genetická a histologická data aj., ale jelikož se toto onemocnění týká konkrétní osoby, často také získáváme spoustu osobních informací o pacientech. K evidenci takovýchto informací, aby mohly být dále zpracovány v nějaký výstup, mohou sloužit databázové systémy.

Databázi se rozumí určitý přesně definovaný soubor dat, sloužící k popisu reálného světa. U astrogliomů to může být evidence pacientů s astroglálními tumory, kde pacient je prvkem z reálného světa, takzvanou *entitou* a tento prvek je popsán svými specifickými vlastnostmi – *atributy* [13], [14]. Obvyklými atributy jsou jméno, bydliště, rodné číslo aj. V souvislosti s onemocněním to může být typ nádoru, jeho stupeň gradingu, nalezené mutace apod.

Prostředkem pro popis databáze je databázový model. Nejčastěji užívaným modelem jsou relační databáze, které si lze představit jako systém tabulek

skládající se ze sloupců a řádků, kde sloupce odpovídají jednotlivým atributům (např. jméno, příjmení, bydliště) a v řádcích je popsán aktuální stav pacienta. V tomto ohledu je velmi důležitou položkou databázového modelu primární klíč, jehož vlastností je právě to, že má jedinečnou hodnotu. Jedinečná hodnota primárního klíče zamezuje situaci, kdy by ku příkladu dva řádky v tabulce mohly mít totožnou hodnotu primárního klíče, což je nežádoucí. U pacientů by tímto primárním klíčem mohlo být rodné číslo. Problémem však je, že v důsledku zákona o ochraně osobních údajů, je tato informace velmi citlivá a nelze ji tudíž použít. Pokud nelze v tabulce použít primární klíč jako určitý sloupec s daty nebo kombinace sloupců s daty, obvykle se v této situaci vytvoří sloupec tzv. id (identifikace), což je sloupec s jedinečnými hodnotami postupně se zvyšujícími čísly. Mnohdy lze vytvořit primární klíč složením všech sloupců, ale takovéto řešení není vhodné, protože zvyšuje náročnost databázového systému na výpočetní výkon. Proto pro tvorbu databáze jsou definovány tzv. normální formy, což je souhrn pravidel, dle kterých by měla být databáze navržena pro optimální využití databázového systému [13]. Mezi tabulkami s konkrétními sloupci jsou relace, které vytváří určitá spojení, jež nalezneme i v reálném světě. Například pacient (jakožto tabulka se sloupci jméno, příjmení, rodné číslo, jméno nádoru) má velký tumor. Tento vztah je definován ve sloupci *nádor* z tabulky pacient pomocí relace na primární klíč sloupce *jméno nádoru* z tabulky *typy nádorů* s konkrétními hodnotami *velký tumor*, *malý tumor*, *drobný tumor*.

Ve světě informačních technologií je pro tvorbu databází nejčastěji používán jazyk SQL (Structured Query Language) a na něm založené databázové systémy. Tento jazyk je nástrojem pro tvorbu databází (tabulek) a zároveň nám umožňuje další manipulaci s daty jako je vkládání dat, aktualizace, mazání a vyhledávání informací aj. [14].

Při navrhování databáze musí být každý krok pečlivě zvolen a zvážen, protože jakákoliv chyba nebo nevhodně zvolený krok může být velmi kritický, kdy tyto chyby v budoucnu mohou způsobit problémy ve funkčnosti databáze, v pozdějších přesu-

nech a úpravách již zadaných dat a podobně.

Další možností zpracování získaných dat jsou statistické metody, kterých existuje celá řada. Výběr statistické metody se odvíjí od toho jaká data a jaké téma potřebujeme zpracovat.

V souvislosti s astroglálními tumory je často sledovaným tématem efektivnost prognostických faktorů na přežívání u pacientů s glioblastomy. Pro odhad střední doby přežití u pacientů se využívá Kaplan-Meierovy metody a účinek proměnné na dobu přežívání je vyhodnocován Log-Rank testem. Proměnnou bývá věk pacienta, pohlaví, Karnofského skóre, lokalizace nádoru v mozku, postoperativní radioterapie a chemoterapie a jiné [15]. Dalšími statistickými metodami, které se hojně využívají v problematice astroglálních tumorů je Chi-kvadrát test, Studentův t-test, Mann-Whitney U test či Spearmanův korelativní koeficient [16].

4. Závěr

Problematika astroglálních tumorů je velmi rozsáhlá a i přes množství provedených studií je spousta otázek, které je potřeba zodpovědět. Nejdůležitějším bodem je hledání nových prognostických a prediktivních faktorů, které by měly význam pro zvolení vhodných terapeutických postupů. Tento problém se nejvíce týká glioblastomů, kde doba přežívání od zjištění diagnózy je velmi nízká. Z množství diagnostikovaných glioblastomů je přibližně 90 % glioblastomů primárních. Tento typ nádoru se vyvíjí velmi rychle, bez předchozí klinické či histologické manifestace. Primární a sekundární glioblastomy se kromě rozdílných mutací liší také profily exprese na úrovni RNA a proteinů a v metylacích promotorových oblastí. Tyto rozdíly zapříčiňují odlišnou vnímavost k radio- a chemoterapii [17], [18]. V současné době se značná část studií zaměřuje na využití mutovaného EGFR receptoru jako molekulárního cíle v nádorové terapii.

Bylo také zjištěno, že metylace O⁶-Methylguanin DNA methyltransferázy (MGMT), klíčového DNA reparačního enzymu, zvyšují citlivost sekundárních glioblastomů k alkylačním činidlům, používaným v chemoterapeutické léčbě [18], [19].

Získání nových a specifitějších informací, týkajících se patogeneze astroglálních tumorů, které by přispěly ke zlepšení a zkvalitnění diagnostiky a léčby těchto tumorů, je dlouhodobým úkolem, který bude vyžadovat ochotnou spolupráci mezi kliniky, patologií a molekulárními biologii.

Práce byla podpořena projektem SVV-2010-265513.

Literatura

- [1] Ehrmann J., Koďoušek R., Kolář Z.: Stručný přehled patologie nádorů nervového systému, LFUP Olomouc; 2007.
- [2] Louis D.N., Ohgaki H., Wiestler O.D., Cavenee W. K.: WHO classification of tumours of the central nervous system. Inter. Agen. for Res. on Can.; 2007.
- [3] Coleman W.B., Tsongalis G.J.: Molecular pathology; The molecular basis of human disease. Elsevier; 2009.
- [4] Ohgaki H. et al.: Genetic pathways to primary and secondary glioblastoma. Amer. Journ. of Pathol. 2007; 170: 1445-1453.
- [5] Puget S. et al.: Malignant brain tumors: Two steps forward. Clin. Neurosur. 2007; 54: 4-9.
- [6] Kolář Z. a kolektiv: Molekulární patologie nádorů. EPAVA Olomouc; 2003.
- [7] Duerr M.E. et al.: PTEN mutations in gliomas and glioneuronal tumors. Oncog. 1998; 16: 2259-2264.
- [8] Aghi K.M. et al.: Decreased rate of infection in glioblastoma patients with allelic loss of chromosome 10q. J. Neurooncol. 2009; 93: 115-120.
- [9] Rasheed Ahmed B.K., Stenzel T.T., et al.: PTEN gene mutations are seen in high-grade but not in low-grade gliomas. Canc. Res. 1997; 57: 4187-4190.
- [10] Wang I. et al.: Somatic mutations of PTEN in glioblastoma multiforme. Canc. Res. 1997; 57: 4183-4186.
- [11] Frederick L. et al.: Diversity and frequency of epidermal growth factor receptor mutations in human glioblastomas. Cancer Research 2000;60: 1383-1387.
- [12] Lukáš Z., Dráberová E. a kol.: Imunohistochemické metody v biologii a v biotické diagnostice. MU Brno; 1997.
- [13] <http://interval.cz/clanky/database-a-jazyk-sql/>
- [14] Oppel A.: SQL bez předchozích znalostí, Comp. Press. 2008; 240.
- [15] Tugcu B., Postalci L. S., Gunaldi O. et al.: Efficacy of clinical prognostic factors on survival in patients with glioblastoma. Turk. Neurosur. 2010; 2: 117-125.
- [16] Martinez R., Rohde V., Schackert G.: Different molecular patterns in glioblastoma multiforme subtypes upon recurrence. J. Neurooncol. 2010; 96: 321-329.
- [17] Voldborg R.B. et al.: Epidermal growth factor receptor (EGFR) and EGFR mutations, function and possible role in clinical trials. A. of Onc. 1997; 8: 1197-1206.
- [18] Kuan C-T, Wikstrand C.J., Bigner D.D.: EGF mutant receptor vIII as a molecular target in cancer therapy. Endocr.-Relat. Can. 2001; 8: 83-96.
- [19] Eoli M., Menghi F. et al.: Methylation of O⁶- methylguanine DNA methyltransferase and loss of heterozygosity on 19q and/or 17p are overlapping features of secondary glioblastomas with prolonged survival. Clin. Can. Res. 2007; 13: 2606-2613.

Kontakt

Mgr. Hana Bielníková
Ústav patologie
Fakultní nemocnice Ostrava
17. listopadu 1790
708 52 Ostrava-Poruba
e-mail: hana.bielnikova@email.cz

Multiagentní architektura založená na procesech aplikovaná na doménu domácí péče

B. Bošanský¹

¹Centrum agentních technologií, Katedra kybernetiky, Fakulta elektrotechnická, České vysoké učení technické v Praze

Školitelka: Doc. Ing. Lenka Lhotská, Csc.

Souhrn

Využití znalostí ve formě postupů, konkrétně organizačních procesů a formalizovaných lékařských doporučení, může být vhodné pro vytvoření znalostní báze systému pro podporu v rozhodování (DSS) v oblasti poskytování zdravotní péče. Problém nastává v případě, že pro vývoj DSS chceme použít multiagentní přístup z důvodu rozdílů mezi formalizací chování agentů a procesním zápisem. V tomto příspěvku pokračujeme v práci na nové multiagentní architektuře a představíme její integraci do stávajícího systému na podporu v rozhodování (K4Care) v oblasti domácí péče.

Základní metodou byla analýza dostupné dokumentace ke komplexnímu systému K4Care, na jejímž základě jsme identifikovali společná místa v rámci již existující funkcionality a návrhu nové architektury. Ty dále posloužily jako výchozí body pro vylepšení modelu K4Care s ohledem na novou multiagentní architekturu založenou na procesech.

Analýza potvrdila nejen možnost takové integrace, ale také její přímocílost a minimum nutných změn v modelu K4Care díky dostatečně obecnému návrhu multiagentní architektury založené na procesech. Na základě integrace byla identifikována okamžitá vylepšení podporující lidského experta při jeho práci se systémem, jakož i možnosti dalšího rozšíření systému K4Care na základě této integrace.

Integrace multiagentní architektury může být přínosná i pro stávající systémy pro podporu v rozhodování a díky ní se otevrou nové možnosti založené na multiagentním přístupu.

Klíčová slova: multiagentní systémy, multiagentní architektury, organizační procesy, formalizované lékařské doporučení, zdravotní péče, domácí péče, K4Care

1. Úvod

Systémy pro podporu v rozhodování (DSS z angl. Decision Support System) si v medicínském prostředí našly své místo a denně poskytují lékařům oporu a usnadňují jim jejich rozhodnutí. Několik studií potvrdilo, že přínos těchto systémů spočívá zejména v předcházení omylů při poskytování zdravotní péče a zvyšování její kvality, ale rovněž vede i ke snížení nákladů a k úspoře času [1-5]. Úspěšnost těchto systémů bývá v praxi podmíněna přístupným a neobtěžujícím ovládáním pro lékaře či ostatní členy lékařského personálu. Toto pozorování je doloženo v práci [2], která systematicky zkoumala stávající DSS, přičemž ty, které pracovaly na pozadí a uživatele upozornili pouze v případě připomínky či varování (příklady takových systémů lze nalézt v [6], [7]), byly vyhodnoceny jako nejvíce přínosné. Znalosti v DSS typicky vycházejí z lékařských doporučení (např. v [8], [9]), které popisují doporučené postupy pro diagnózu a/nebo postup léčby konkrétních onemocnění. Z analýzy existujících DSS [1-9] však rovněž vyplývá, že pro jejich úspěšné použití je nutná plná integrace v daném zařízení, kde jsou používány, což znamená přizpůsobení lokálním podmínkám a omezením. Proto je nutné, aby se kromě obecných lékařských doporučení při tvorbě znalostní báze DSS zohledňovaly také organizační postupy v daném zařízení. Tyto znalosti zachycující medicínské postupy pak mohou být systémem použity pro kontrolu jejich návrhu, simulaci nebo na kontrolování jejich dodržování v praxi.

Jednou z oblastí medicíny, kde nasazení DSS může být mimořádně přínosné, je oblast domácí péče (HC z angl. Home care) [10], [11]. Zvyšující se počet chronicky nemocných starších pacientů má za následek zvýšené nároky na nepřetržitou dlouhodobou léčbu pod dohledem zdravotních specialistů. Navíc

je možné pozorovat, že umístění a léčba těchto pacientů v domácím prostředí může mít pozitivnější výsledky než jejich dlouhodobá hospitalizace [11]. Potřeba nasazení DSS v oblasti domácí péče vychází zejména z její distribuované povahy. Pracovníci domácí péče navštěvují doma své pacienty a poskytují jim patřičnou asistenci či vyšetření, takto však přicházejí o možnost rychlé konzultace se svými kolegy při neznámých situacích nebo komplikacích. Nasazení systému pro podporu nebo kontrolu jejich činností by tak mohlo jednoznačně zvýšit kvalitu poskytované zdravotní péče.

Vzhledem k distribuované povaze oblasti domácí péče ji musí zohledňovat také příslušný DSS. Z tohoto důvodu jsou často využívány modely vycházející z multiagentních přístupů [11-15]. Pod pojmem multiagentní systém rozumíme systém, kde skupina inteligentních autonomních agentů společně interaguje, přičemž chtějí dosáhnout předem definované cíle. Chování těchto agentů bývá specifikováno pomocí více základních modelů – (1) čistě reaktivní agenti, kteří na základě předdefinovaných pravidel reagují na podněty okolí, (2) uvažující agenti, kteří si budují vlastní model prostředí a plánují své budoucí akce a (3) hybridní modely kombinující předchozí dva modely. Tyto základní modely chování agentů mohou být implementovány jak s využitím standardních programovacích jazyků, např. Java, tak s využitím specializovaných agentních programovacích jazyků, jako např. Jazzyk [16], CONRAD [17], nebo některé z verzí jazyka APL [18]. Nedostatkem těchto programovacích jazyků je zejména nemožnost efektivně namodelovat situace kooperace a koordinace tak, aby agenti sledovali vlastní cíle, ale zároveň i cíle společné, pro které je kooperace nutná.

Na druhou stranu, organizační procesy a lékařská doporučení jsou často založeny na nutnosti kooperace zúčastněných aktérů, přičemž poskytují i potřebné informace o následnosti a podmínkách provedení jednotlivých akcí (např. před operačním zákrokem je nutno provést příslušná vyšetření, pro která je nutný jistý druh vybavení apod.). V předchozích pracích jsme proto vytvořili nový způsob modelování chování agentů, který takovou kooperaci podporuje [19], [20]. Náš přístup je založen na rozšířených procesech, pomocí kterých formalizujeme znalosti o postupech. Na jejich základě vytváříme agenty systému, jejichž specifická organizace dovoluje pracovat s procesy ve smyslu kontroly jejich návrhu, jejich simulace, případně kontroly jejich exekuce v praxi. Z důvodu přehlednosti budeme v této práci tuto architekturu označovat jako ProMA.

Architektura ProMA byla původně navržena s ohledem na všeobecné použití a tedy s obecnou specifikací procesního formalismu. Cílem této práce je proto ukázat možné aplikace architektury ProMA pomocí rozšíření architektury stávajícího systému pro podporu domácí péče, K4Care [11], kde se pro ukládání postupů využívá formalismus SDA* [21]. Důraz při tom klademe na integraci nové architektury do systému K4Care a její dopad na změny v modelech a nové vlastnosti takto vylepšeného systému. Příspěvek je organizován tímto způsobem. V následující sekci přesněji vysvětlíme pojem procesu v medicínském kontextu, poté následuje popis nové architektury ProMA, která s procesy dokáže pracovat. V další sekci představíme systém K4Care, po ní pak následuje sekce věnovaná zapojení architektury ProMA do systému K4Care. Zdůrazníme, že toto zapojení slouží jako potvrzení aplikovatelnosti obecné architektury ProMA v konkrétním systému, přičemž popíšeme zejména benefity plynoucí z této integrace. V předposlední sekci nastíníme nové možnosti, které tato integrace poskytuje, a v poslední sekci poskytneme shrnující závěr.

2. Multiagentní architektura založená na procesech (ProMA)

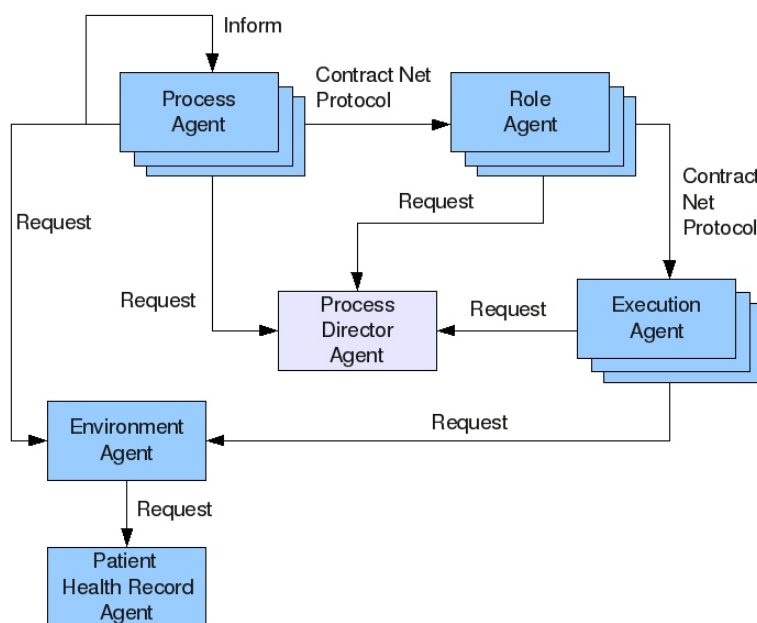
Tato sekce je věnována popisu procesů v oblasti zdravotnictví, jejich rozdělení a následně architektury ProMA, která s nimi umožňuje pracovat. Pod pojmem proces rozumíme sekvenci aktivit, stavů, rozhodnutí, ale také stavů rozdělovajících sekvenci na několik paralelních, resp. synchronizačních stavů. Procesy v oblasti medicíny rozdělujeme na dva základní typy – *organizační procesy* a *procesy související s léčbou* (resp. *lékařská doporučení*).

Organizační procesy v oblasti zdravotní péče jsou podobné procesům známým z jiných odvětví. Několik prací (např. [22], [23]) se věnuje zkoumání možností aplikace procesního modelování či používání tzv. workflow management systémů v oblasti zdravotnictví. Panuje shoda, že úspěšné nasazení těchto technik a systémů může přinést zvýšenou kvalitu poskytované péče, snížit čas potřebné hospitalizace pacienta a díky tomu i snížit celkové náklady.

Lékařská doporučení jsou delší dobu součástí procesu standardizace lékařské péče a lze je chápat jako jeden z pilířů medicíny založené na znalostech.

Obsahují doporučené a odbornými komisemi schválené postupy, akce a principy pro diagnostiku či léčbu onemocnění. V současné době mají používaná lékařská doporučení často textovou podobu, která je pro lékařské specialisty přirozená. Na druhé straně je však tato textová podoba komplikací při potřebě rychlé konzultace během vyšetření pacienta, případně pro rychlé zorientování se ve změnách při vydání nové varianty doporučení. Navíc, textová podoba je naprosto nevhodná pro počítačové zpracování, nebo jako vědomostní základ pro systémy pro podporu v rozhodování. Proto je velká část výzkumu v oblasti medicínské informatiky věnována právě formalizaci lékařských doporučení do elektronické podoby, přičemž bylo vyvinuto několik procesních jazyků pro takovou formalizaci (např. jazyky PROforma [24], Asbru [25] nebo GLIF [26]) a s jejich pomocí lze zachytit znalosti z textových doporučení do strukturované elektronické podoby.

Oba tyto typy procesů můžeme vnímat společně a oba musí být zastoupeny při tvorbě báze znalostí postupů v DSS. V další části této sekce představíme multiagentní architekturu, která je s těmito znalostmi formalizovanými pomocí obecného procesního jazyka schopna pracovat. Celkový náhled na architekturu je na Obrázku 1.



Obr. 1. Schéma multiagentní architektury založené na procesech.

2.1 Agent prostředí (Environment Agent)

Multi-agentní systémy a simulace jsou vždy vztaženy k určitému prostředí, ve kterém agenti operují. V ProMA architektuře je toto prostředí reprezentováno specializovaným agentem. V závislosti na tom, na jaké úrovni detailu chceme systém používat, může tento agent reprezentovat virtuální svět (to znamená od celého nemocničního oddělení po položky v databázi záznamů pacientů) spolu se stávajícími objekty (např. lůžka, RTG nebo EEG přístroje, laboratorní výsledky apod.). Je však rozdíl při práci s daty, které se týkají zdravotního záznamu pacientů a ostatními položkami. Proto, jak je uvedeno ve schématu, a jak bude ukázáno později, je vhodné oddělit simulační a skutečná data, avšak, agent prostředí je schopen tyto údaje získat od specializovaného agenta (Patient Health Record Agent).

2.2 Exekuční agent (Execution Agent)

Exekuční agenti (EA) v architektuře reprezentují konkrétní lidi – lékaře, sestry, pacienty či dalších zaměstnance, kteří se podílejí na výkonu formalizovaných procesů. Tito agenti využívají reaktivní model chování formou hierarchických pravidel, které lze na základě procesní specifikace automaticky generovat (v publikaci [20] jsou uvedeny detaily daného algoritmu). Každý EA má předem definovanou sadu pravidel, pomocí kterých je schopen plnit základní cíle v daném prostředí (např. odpovídá na zaslané zprávy, informuje agenta prostředí o nutných změnách stavu apod.). Následně, pro každou aktivitu, ve které může tento agent (a tedy konkrétní osoba, kterou reprezentuje) vystupovat, je vytvořeno jedno další pravidlo. Tyto mohou být v průběhu práce s procesy aktivovány (v případě, že je možné danou aktivitu uskutečnit) nebo deaktivovány (v případě, že provedení této aktivity není aktuálně možné). Aktivaci či deaktivaci provádí agent role (bude popsán později). Výhodou tohoto systému je fakt, že to, kterou konkrétní aktivitu bude exekuční agent provádět, závisí na jeho autonomním rozhodnutí a jeho prioritách k aktivovaným prioritám.

2.3 Agent role (Role Agent)

Agenti rolí (RA) reprezentují všeobecné role v daném prostředí (všechny lékaře, sestry atd.). Cílem RA je najít vhodného exekučního agenta (EA) pro vykonání dané aktivity na žádost příslušného Procesního Agentu (bude popsán později). Důvodem, pro který je vhodné nasazení specifických agentů reprezentujících role je fakt, že v typické pracovní struktuře zaměstnanců jsou jejich role navzájem hierarchicky uspořádány (např. sekretářka i sestra jsou obě zároveň i zaměstnankyněmi). Proto, když RA obdrží žádost o účast na dané aktivitě od procesního agenta, vyhledává mezi exekučními agenty (pomocí tzv. contract-net protokolu (CNP)) ty, kteří mají tuto roli, ale i ty role, které jsou v kontextu dané hierarchie všeobecnější.

V případě, že jednu aktivitu může současně provést několik EA a je potřebný jen jeden, RA autonomně určí vítěze na základě interních pravidel, přičemž vítězi předá zprávu o aktivaci daného pravidla. Výběr takového vítěze je vždy doménově závislý, případně závisí na konkrétních rolích (např. se může jednat o agenta, který byl nejdéle nečinný apod.). Navíc, RA jsou odpovědní i za nalezení náhrady za vítězného EA v případě, že ten dočasně pozastaví činnost na dané aktivitě (např. lékař je zavolán k akutnímu případu a může být nahrazen v aktuální činnosti).

2.4 Procesní agent (Process Agent)

Procesní agent (PA) je vytvořen pro každý krok v procesním formalismu (tedy pro každý stav, aktivitu, rozhodnutí, rozdělení či synchronizaci). PA je zodpovědný za korektní exekuci daného kroku. Na začátku kontroluje zda jsou splněny vstupní podmínky pro zahájení procesu:

- zda předchozí PA úspěšně skončil(i) (a tedy zda PA obdržel příslušnou informační zprávu,
- zda všechny hodnoty na objektech potřebných k provedení tohoto kroku jsou správné (PA se pomocí jednoduchého protokolu ptá agenta prostředí),
- zda existují exekuční agenti, kteří jsou potřební pro provedení tohoto kroku (PA používá CNP protokol pro příslušné agenty rolí, kteří přísluší této aktivitě).

V případě, že jsou všechny nutné podmínky splněny, zahájí PA exekuci kroku (tedy např. simulaci aktivity, výpočet nebo učiní patřičné rozhodnutí) a po úspěšném ukončení je PA povinen zaslat výsledek agentovi prostředí (pomocí jednoduchého protokolu žádosti) a informovat vlastního následovníka o úspěšném konci (pomocí jednoduchého informačního protokolu).

Tato architektura navíc umožňuje dočasně pozastavit provádění patřičné aktivity a promítnutí dílčích výsledků do prostředí, nahrazení jednoho exekučního agenta druhým, koordinaci více exekučních agentů i nepovinné vstupní podmínky.

2.5 Agent Procesního Záznamu (Process Director Agent)

Poslední agent je výhradně pomocným agentem, který čte formalizované procesy, na základě kterých vytváří ostatní agenty a následně jim odpovídá na otázky týkající se těchto procesů (jako např. vstupní objekty, předchůdci, apod.). Jedná se o zjednodušení vlastní implementace systému postaveném na této architektuře, ale je vhodné upozornit, že stále všichni agenti jsou si vědomi a uvažují o procesech jako takových.

3. K4Care

Cílem projektu K4Care [11] bylo vytvoření, implementace a ověření zdravotního modelu postaveného na znalostech pro asistenci pacientům v domácí péči. Znalosti o postupech jsou reprezentovány pomocí lékařských doporučení, která se v kontextu tohoto projektu nazývají formální intervenční plány (FIP). Na druhou stranu, pro zachycení specifických podmínek pro každého pacienta představuje projekt K4Care tzv. individuální intervenční plány (IIP), které spojují jak medicínské znalosti, tak i lokální pravidla a sociální charakteristiky daného pacienta.

Znalost o obou typech postupů, tedy jak individuálních tak i formálních intervenčních plánech, je v systému K4Care reprezentována pomocí formalismu SDA* [21]. Jedná se o procesní jazyk obsahující 3 základní terminologické komponenty: (1) stavy, (2) rozhodnutí a (3) akce. Stavy popisují přípustné stavy pacienta nebo významné situace v daném kontextu (např. zvýšený krevní tlak).

Rozhodnutí reprezentují terminologii, kterou specialisté používají pro rozlišení několika následujících postupů (např. dřívější problémy se srdcem mohou rozlišit následující kroky, pokud jsou pravdivé). Termíny akcí popisují lékařské, klinické nebo operační akce, které mohou provádět aktéři zohlednění v systému (např. užití betablokátoru, vyhýbání se slaným jídlům, provedení krevní analýzy, apod.). Na rozdíl od typických procesních jazyků, SDA* nemá specifickou komponentu pro aktéry, ale ti jsou v systému fixně určeni ke každé akci. Rozlišujeme dvě skupiny — *žadatelé* a *vykonavatelé*. V první skupině jsou ti aktéři, kteří mají oprávnění žádat provedení dané akce, ve druhé skupině jsou ti aktéři, kteří mají oprávnění k jejímu provedení v praxi.

Formalizované znalosti postupů jsou nejprve využívány lidským expertem (lékaři nebo manažeri daného zdravotního zařízení). V případě, že pacient prostřednictvím systému požádá o poskytnutí služby, tým specialistů nejprve identifikuje na základě příznaků množinu možných onemocnění. Platforma K4Care následně poskytne dostupné formálně intervenční plány (FIP), které se daných znalostí mohou dotýkat.

Lidský expert následně spojí a dále upraví tuto množinu plánů tak, aby odpovídaly aktuálním potřebám. Výsledný plán je uložen jako individuální intervenční plán daného pacienta. Tento plán má přiřazené konkrétní aktéry, kteří se budou na provádění tohoto plánu podílet, a proto se po jejich odsouhlasení výsledné úkoly zobrazí v jejich osobních portálech. Tyto jsou vygenerovány SDA* exekučním mechanismem na základě uloženého IIP. Pracovník po úspěšném vykonání dané úlohy vyplní v systému příslušná místa v dokumentu týkající se dané aktivity a vygenerují se nové úkoly na základě specifikace v IIP.

4. Integrace architektury ProMA do systému K4Care

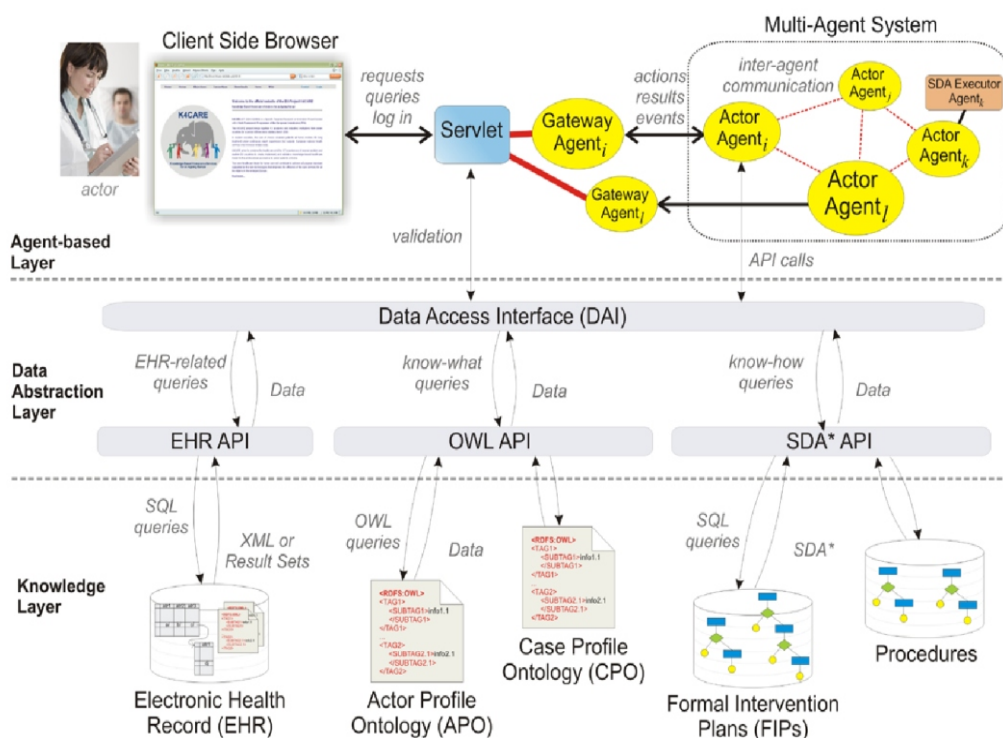
V této sekci popíšeme vylepšení stávajícího modelu systému K4Care integrací architektury ProMA. Nejdříve se zaměříme na technické detaily vyplývající z této integrace, pak bude následovat návrh nových možností, které takto modifikovaný systém může poskytnout. V celé sekci následujeme příklad použití uvedený v závěru předchozí sekce a integraci budeme vysvětlovat v tomto kontextu. Důvodem je přímo viditelná výhoda při návrhu IIP lidským expertem díky integraci architektury ProMA do systému K4Care.

Další možnosti rozšíření a z nich vyplývající nové vlastnosti systému diskutujeme v následující sekci.

4.1 Prostředí

V tomto příspěvku se zaměřujeme na multiagentní část platformy K4Care (viz Obr. 2). Nejdříve se proto zaměříme na unifikaci stávající funkcionality systému K4Care s architekturou ProMA. Propojení s údaji ve zdravotních záznamech pacienta a doporučenými postupy v architektuře ProMA reprezentované pomocí komunikace příslušných agentů, je v systému K4Care realizováno pomocí datové abstrakční vrstvy (DAL). Ta poskytuje rozhraní (tzv. API z angl. Application Program Interface) pro získání dat pacientů z jejich záznamu nebo doporučení formalizované pomocí SDA*.

Přesto, agent prostředí musí být stále přítomen pro dočasné ukládání dat, která nemají perzistentní charakter (např. během simulace). Vrstva DAL rovněž neumožňuje odpovědět na dílčí dotazy pro jednotlivé položky dokumentů či doporučení, proto oba specializovaní agenti (agent procesního záznamu a agent pacientova zdravotního záznamu) musí být ve výsledném multiagentním systému vytvořeni.



Obr. 2. Schéma systému K4Care prezentovaná v [13].

4.2 Exekuční agenti

Exekuční agenti (EA) architektury ProMA konceptuálně odpovídají agentům aktérů v systému K4Care (Actor Agents) - reprezentují totiž konkrétní lidi, uživatele systému, kteří vykonávají akce. Pro úplnou shodu však agenti aktérů musí umět komunikovat s agenty rolí (RA) a umět autonomně reagovat na jejich požadavky. Taková reakce je v současném návrhu systému K4Care postavená výhradně na lidské odpovědi. Vzhledem k tomu, že jsou tito agenti aktérů perzistentně v systému přítomní, jejich funkcionalita může být poměrně jednoduše vylepšená tak, aby dokázaly reagovat na příchozí požadavky v podobě CNP protokolů inicializovaných agenty RA. Autonomní odpověď může být postavena na seznamu aktuálně přidělených úkolů danému aktérovi nebo jeho aktuální dostupnosti (pokud je například specialista dlouhodobě mimo zařízení, jeho agent by měl mít tuto informaci a automaticky zamítnout jeho účast na navrhovaných aktivitách). Navíc, koncepty formalismu SDA* typicky obsahují i časové informace (např. časový interval či periodicitu výskytu), díky čemuž mohou poskytovat agenti aktérů mnohem sofistikovanější odpovědi v závislosti na všech plánovaných aktivitách a přidružených IIP, přičemž ve finále mohou společně fungovat jako rozvrhovací algoritmus umožňující lidskému expertovi jednodušší rozhodnutí při vytváření nového IIP (více je tato alternativa analyzována v následující sekci).

4.3 Agenti rolí

Ostatní typy agentů prezentované v architektuře ProMA nemají přímou alternativu v systému K4Care, a proto musí být vytvořeni, přičemž v dalších sekcích popíšeme jejich integraci do systému. Jak agenti rolí (RA), tak ani procesní agenti (PA) nemusí být v systému perzistentně přítomní a mohou být vytvořeni při vytváření nového IIP pro daného pacienta na základě vyžádaných doporučení (FIP).

Agenti rolí (RA) konceptuálně v systému K4Care odpovídají typům aktérů. Proto vytvoříme RA pro každou následující roli: pacient, rodinný lékař, hlavní sestra, vedoucí lékař, sociální pracovník, sestra, lékař specialista, operátor sociálního pracoviště, poskytovatel nepřetržité péče a ostatní. V systému K4Care jsou navíc

tyto role sdruženy do skupin jako evaluační jednotka či další poskytovatelé péče. Proto jsou RA vytvořeni i pro každou z těchto skupin a příslušná hierarchická struktura mezi rolemi je vybudována. Navíc, ve více aktivitách také existuje podobná hierarchie (např. řadu aktivit, které provádí sestra umí provést i hlavní sestra, apod.).

4.4 Procesní agenti

Jeden procesní agent (PA) je vytvořen pro každý stav, rozhodnutí nebo aktivitu ve vyžádaných doporučeních FIP formalizovaných pomocí SDA*. Také, pokud uživatel v IIP vytvoří nové kroky, pro každý z nich se vytvoří nový PA. Kolektivní chování v základním režimu odpovídá jedinému agentovi v systému K4Care – SDA* exekučnímu mechanismu. Výhody oproti stávajícímu modelu jsou viditelné například v částečném automatickém spojení FIP, který může navrhnout expertovi základní variantu finálního IIP, kterou on/ona následně upraví podle vlastních znalostí a zkušeností.

Díky tomu, že všechny charakteristiky kroků formalizovaných pomocí SDA* využívají zápis pomocí rozsáhlé ontologické báze, můžeme tyto informace využít pro částečnou automatizaci procesu spojení vyžádaných FIP. Ekvivalence několika kroků modelu SDA* může být provedena pomocí příslušných PA, kteří si porovnávají asociované termíny, kontext a vstupní podmínky. V případě, že souhlasí, ve finálním IIP může být umístěn jen jeden PA, který využívá sjednocení všech vstupních a výstupních údajů (např. krevní vyšetření může být provedeno pouze jednou pro všechny požadované informace). Myšlenka poloautomatického spojování dostupných FIP není nová (základní návrh byl prezentován v [13]), ale návrh v tomto příspěvku se zaměřuje na spojování jednotlivých kroků postupů pomocí vyjednávání autonomních agentů. Tato základní myšlenka může být dále vylepšena pomocí základní simulace (bude vysvětleno později) všech otevřených postupů (což architektura ProMA umožňuje). Jednotliví PA takto umí identifikovat pořadí, ve kterém mohou být exeklováni nejen v závislosti na pořadí v jednotlivých FIP, ale také v závislosti na splnění dalších vstupních podmínek. Pod pojmem „základní simulace“ myslíme simulaci, kdy se PA snaží vyplňovat

potřebné výstupní hodnoty dočasnými libovolnými hodnotami a také díky tomu analyzovat možný další postup v jednotlivých FIP. V následující sekci tuto myšlenku dále rozšíříme.

Nakonec, pomocí CNP protokolu delegují PA problém výběru aktéra na asociované agenty rolí (RA). Oproti původnímu návrhu v architektuře ProMA musí RA v K4Care dodržovat bezpečnostní oprávnění a tedy notifikovat pouze ty agenty aktérů, kteří mají oprávnění provádět tuto akci pro pacienta, pro kterého je aktuálně IIP vytvářeno. Na základě odpovědi agentů aktérů může RA vrátit PA nejlepšího kandidáta, ale také můžeme tuto funkcionalitu modifikovat a RA vrátí seřazený seznam kandidátů a výběr přenechá expertovi.

5. Možné vylepšení a diskuse

V předchozí sekci jsme popsali integraci architektury ProMA do systému K4Care, která potřebovala jen minimální počet konceptuálních změn v daných modelech. Klíčová výhoda takové integrace však spočívá v otevření nových možností pro systém K4Care, které jsou díky architektuře ProMA snadněji proveditelné. Na druhou stranu by vyžadovaly rozšíření znalostí o aktivitách podporovaných tímto systémem.

Klíčovým faktorem je využití simulace všech postupů během tvorby nového IIP. Výhodou systému K4Care je, že oba typy procesů, jak lékařská doporučení, tak i organizační procesy, jsou formalizované pomocí stejného jazyka SDA* a jsou tak vnímány a je s nimi nakládáno na ekvivalentní úrovni. Jak bylo ukázáno v předchozí sekci, i základní simulace je schopna poskytnout základní omezení a návrhy, jejichž vizualizace může expertovi usnadnit práci při návrhu nového IIP.

Předpokládáme tedy dále, že v systému K4Care budou obsaženy doplňující informace. Za prvé rozšíříme stávající ontologický systém o další časové informace jako odhadovaná délka trvání jednotlivých aktivit a specifikace postupu pomocí časové funkce (viz [20] pro detailnější popis těchto funkcí).

Dále předpokládejme, že součástí prostředí je mapa, resp. virtuální prostředí ve kterém dané zařízení působí. V poslední řadě předpokládáme povolení přístupu PA k datům ostatních pacientů (samozřejmě omezených dle bezpečnostních pravidel). Všechny tyto dodatečné informace nám umožní provést simulaci přesněji s důrazem na potřebné detaily. V závěru tvorby IIP může být tento návrh simulován a ohodnocen podle přiřazených aktérů či dalších kritérií.

Simulace postupů pracuje v souladu s návrhem architektury ProMA, což znamená, že PA se snaží úspěšně simulovat asociovaný krok postupu výběrem vhodných agentů rolí a následně agentů aktérů. Vzhledem k tomu, že agenti aktérů jsou limitováni dalšími již zadanými úkoly (to znamená, že vždy simulujeme i ostatní aktuálně platné IIP uložené v systému), vybírají si akce, na kterých budou pracovat na základě vlastních možností a priorit. Tuto myšlenku můžeme demonstrovat na jednoduchém příkladu: potřebujeme přiřadit sestru k aktivitě N.15 – změření hladiny cukru v krvi doma u pacienta. V systému je k dispozici sestra Alice a hlavní sestra Eva. Alice má naplánovanou jednu úlohu na opačném konci města, Eva má plánované dvě kratší úkoly v centru. Přestože aktérky Alice a Eva mají různé role, každá z nich může tuto aktivitu vykonávat díky povolené hierarchii (kdy sestra je obecnější role jako hlavní sestra). Díky tomu, že architektura ProMA tuto hierarchii umožňuje zvládnout, RA odpovídající roli sestry osloví s požadavkem i RA odpovídající roli hlavní sestry. Pro získání vhodné odpovědi simulujeme celkový stav ve virtuálním prostředí. To znamená, že agenti aktérů v tomto prostředí také simulují provádění již naplánovaných aktivit. Když obdrží nový požadavek na provedení aktivity N.15 od RA odpovídajícího roli sestry, nacházejí se tiho aktéři na různých místech ve virtuálním prostředí a na základě aktuálního stavu ostatních úkolů mohou odhadnout čas, potřebný k splnění daného požadavku. Proto v našem zjednodušeném příkladu, by byla hlavní sestra Eva navržena jako vhodnější kandidát pro vykonání dané aktivity. Pokud by však tato aktivita N.15 měla pro Alici vyšší prioritu a ona by mohla dočasně přerušit provádění právě přiděleného úkolu, mohla by být vybrána

právě ona. Pro věrnou simulaci a zejména výsledky jednotlivých kroků je nutné, aby měli PA přístup k záznamům o pacientech a uměli lépe odhadnout výsledek daného kroku (podobně jako v případě verifikace v [19]). Tento výsledek může totiž významně ovlivnit další pokračování v simulovaných IIP. Navíc, po vytvoření a uložení IIP do systému, může architektura ProMA stejně posloužit pro kontrolu dodržování těchto postupů a v případě inkonzistence upozornit vedoucího pracovníka (viz [19]).

6. Závěr

V tomto příspěvku jsme analyzovali použitelnost obecné multiagentní architektury do stávajícího systému pro podporu v rozhodování, identifikovali jsme potřebné změny v daných modelech a ukázali, jak okamžité výhody z této integrace, tak i potenciální výhody a nové možnosti. Konkrétně jsme popsali procesy a stávající formalismy pro zachycení znalostních postupů v oblasti zdravotnictví a také novou multiagentní architekturu (ProMA), která je s těmito procesy schopna pracovat. Následně jsme se zaměřili na systém K4Care, kde jsou znalostní postupy formalizovány pomocí jazyka SDA* a používané multiagentním systémem pro poskytování podpory v oblasti domácí péče. Navrhli jsme spojení těchto dvou architektur a identifikovali minoritní změny v obou modelech, přičemž jsme poukázali na nové možnosti zefektivňující tvorbu individuálních intervenčních plánů pacienta. Na závěr jsme představili další nové možnosti, které touto integrací vznikly a vlastnosti, které by to systému K4Care mohlo přinést.

Poděkování

Poděkování patří doc. Ing. L. Lhotské, CSc. za poskytnutí materiálů k systému K4Care a Mgr. A. Horňákové za překlad ze slovenského do českého jazyka. Tento výstup vznikl v rámci podpory projektem SVV-2010-265513 a projektem MŠMT ČR č. MSM6840770038.

Literatura

- [1] Balas E.A., Weingarten S., Garb C.T., Blumenthal D., Boren S.A., Brown G.D.: Improving preventive care by prompting physicians. *Arch Intern Med.* 2000 Feb 14;160(3):301308.
- [2] Garg A.X., Adhikari N.K., McDonald H. et al.: Effects of computerized clinical decision support systems on practitioner performance and patient outcomes: a systematic review. *JAMA.* 2005 Mar 9;293(10):12231238.
- [3] Grimshaw J.M., Russell I.T.: Effect of clinical guidelines on medical practice: a systematic review of rigorous evaluations. *Lancet.* 1993 Nov 27;342(8883):13171322.
- [4] Hunt D.L., Haynes R.B., Hanna S.E., Smith K.: Effects of computer-based clinical decision support systems on physician performance and patient outcomes: a systematic review. *JAMA.* 1998 Oct 21;280(15):13391346.
- [5] Kawamoto K., Houlihan C.A., Balas E.A., Lobach D.F.: Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success. *Bmj.* 2005 Apr 2;330(7494):765.
- [6] Burack R.C., Gimotty P.A.: Promoting screening mammography in inner-city settings: the sustained effectiveness of computerized reminders in a randomized controlled trial. *Med Care.* 1997;35:921-931.
- [7] Demakis J.G., Beauchamp C., Cull W.L. et al.: Improving residents' compliance with standards of ambulatory care: results from the VA Cooperative Study on Computerized Reminders. *JAMA.* 2000;284:1411-1416.
- [8] Overhage J.M., Tierney W.M., McDonald C.J.: Computer reminders to implement preventive care guidelines for hospitalized patients. *Arch Intern Med.* 1996;156:1551-1556.
- [9] Nilasena D.S., Lincoln M.J.: A computer-generated reminder system improves physician compliance with diabetes preventive care guidelines. *Proc Annu Symp Comput Appl Med Care.*
- [10] Egan M., Wells J., Byrne K., Jaglal S., Stolee P., Chesworth B. et al.: The process of decision-making in home-care case management: implications for the introduction of universal assessment and information technology. *Health & Social Care in the Community.* (2009, July); 17(4): 371-378.
- [11] Campana F., Moreno A., Riano D., Varga L.Z.: K4care: Knowledge-based homecare e-services for an ageing europe. *Agent Technology and e-Health.* 2008;p. 95115.
- [12] Aubrecht P., Lhotska L., Dolezel J., Dolezel J.: Mobile Devices for e-Services in Home Care. In: 4th European Conference of the International Federation for Medical and Biological Engineering. Springer; 2009. p. 10061009.

- [13] Isern D., Moreno A., Sanchez D., Hajnal A., Pedone G., Varga L.: Agent-based execution of personalised home care treatments. *Applied Intelligence*. 2009;p. 126.
- [14] Isern D., Moreno A.: Computer-based execution of clinical guidelines: A review. *International Journal of Medical Informatics*. 2008;77(12):787–808.
- [15] Isern D., Sanchez D., Moreno A.: HeCaSe2: A Multi-agent Ontology-Driven a Guideline Enactment Engine. In: CEEMAS '07: Proceedings of the 5th international Central and Eastern European conference on Multi-Agent Systems and Applications V. Berlin, Heidelberg: Springer-Verlag; 2007. p. 322324.
- [16] Novak P.: Jazyk: A programming language for hybrid agents with heterogeneous knowledge representations. *Programming Multi-Agent Systems*. 2009;p. 7287.
- [17] Bordini R.H., Hubner J.F., Wooldridge M.J.: Programming multi-agent systems in AgentSpeak using Jason. Wiley-Interscience; 2007.
- [18] Hindriks K.V., De Boer F.S., Van der Hoek W., Meyer J.J.C.: Agent programming in 3APL. *Autonomous Agents and Multi-Agent Systems*. 1999;2(4):357401.
- [19] Bosansky B., Lhotska L.: Agent-based process-critiquing decision support system. In: 2nd International Symposium on Applied Sciences in Biomedical and Communication Technologies, 2009. ISABEL 2009; 2009. p. 16.
- [20] Bosansky B., Brom C.: Agent-Based Simulation of Business Processes in a Virtual World. In: Corchado E, Abraham A, Pedrycz W, editors. *Hybrid Artificial Intelligence Systems*. vol. 5271 of Lecture Notes in Computer Science. Springer Berlin / Heidelberg; 2008. p. 8694.
- [21] Riano D.: The SDA* Model: A Set Theory Approach. CBMS 2007, Maribor, Slovenia, (2007).
- [22] Lenz R., Blaser R., Beyer M., Heger O., Biber C., Baumlein M. et al.: IT support for clinical pathwaysLessons learned. *International Journal of Medical Informatics*. 2007;76(Supplement 3):S397–S402. Ubiquity: Technologies for Better Health in Aging Societies - MIE 2006.
- [23] Kumar A., Smith B., Pisanelli M., Gangemi A., Stefanelli M.: Clinical Guidelines as Plans: An Ontological Theory. *Methods of Information in Medicine*. 2006;2.
- [24] Fox J., Johns N., Rahmzadeh A., Thomson R. PROforma: A method and language for specifying clinical guidelines and protocols. In: Amsterdam; 1996.
- [25] Shahar Y., Miksch S., Johnson P.: The Asgaard project: a task-specific framework for the application and critiquing of time-oriented clinical guidelines. *Artificial Intelligence in Medicine*. 1998;14:2951.
- [26] Peleg M., Boxwala A., Ogunyemi O.: GLIF3: The Evolution of a Guideline Representation Format. *Proc AMIA Annu Fall Symp*. 2000;p. 645649.

Kontakt

Mgr. Branislav Bošanský

Centrum agentních technologií

Katedra kybernetiky

Fakulta elektrotechnická

České vysoké učení technické v Praze
Technická 2

166 27 Praha 6

e-mail: bosansky@labe.felk.cvut.cz

3D-CT rekonstrukce náhrady temporomandibulárního kloubu před a po terapii

E. Feltlová¹, T. Dostálová¹, M. Kašparová¹, J. Daněk², P. Hlíňáková¹, M. Hubáček¹, J. Nedoma¹

¹Dětská stomatologická klinika FN Motol, 2.LF UK v Praze

²Oddělení numeriky katedry matematiky, Fakulta aplikovaných věd ZČU v Plzni, Plzeň

Školitelka: Prof. MUDr. Taťjana Dostálová, DrSc., MBA

Souhrn

V dnešní době moderní pokrok v biomedicině umožňuje lépe pochopit funkci temporomandibulárního kloubu (TMK) a to pomocí zobrazovacích metod. Zpracování získaných dat pomocí moderních programů vytváří nový pohled na rekonstrukci TMK, dále usnadňuje diagnostickou a terapeutickou fázi. K zobrazení TMK jsou k dispozici různé zobrazovací metody. Mezi ně patří základní rtg snímky (např. Ortopantomogram, tomogramy TMK), ultrasonografie, MRI, spirální a 3D-počítačová tomografie. Na základě těchto metod lze snímat požadované úseky čelistního kloubu v různých rovinách a ze získaných dat vytvořit model TMK ve 3D rekonstrukci. To přináší zkvalitnění diagnostického procesu. Cílem naší studie je ukázka využití zobrazovacích metod při sledování pacientky během 1 roku.

V této studii se zabýváme analýzou matematického a počítačového 3D modelu TMK. U pacientky sledujeme léčebný plán krok za krokem od konzervativní terapie, chirurgické terapie TMK, až k procesu hojení. 3D-CT modelace nám umožňuje sledovat jednotlivé fáze léčby u pacientky s radikální cystou dolní čelisti.

Zpracováním matematického a počítačového modelu TMK do 3D-CT modelu TMK analyzujeme stav kloubu před chirurgickým zákrokem a rok po zavedení subtotální náhrady TMK. Modelace zobrazuje stabilitu subtotální náhrady TMK.

Matematické modelování a matematická simulace náhrady čelisti a/nebo TMK může být využito v plánování zhotovení kloubní náhrady. Tato studie prokazuje praktické využití 3D-CT snímků při terapii TMK. Další aplikace matematického modelování čelisti může být využito k zobrazení změn po ortognátní operaci, k hodnocení vlivu okluzních dlah na zatížení TMK, stejně jako k hodnocení u nádorových stavů

a cyst. Ačkoliv mnoho z poruch čelistního kloubu a souvisejících struktur může být diagnostikováno klinicky na základě anamnézy a fyzikálního vyšetření, je mnoho poruch TMK, které můžeme diagnostikovat jen na základě zobrazovací techniky pro přesnou diagnózu a určení rozsahu postižení. Zobrazení tvrdých a měkkých tkání čelistního kloubu může být použito nejen pro léčbu, plánování, ale také pro hodnocení dlouhodobé léčby.

Klíčová slova: stomatologie, maxilofaciální chirurgie, subtotální náhrada TMK, 3D-CT, magnetická rezonance, temporomandibulární kloub

1. Úvod

Čelistní kloub (TMK) je párový složený kloub, který tvoří společné připojení dolní čelisti k spánkové kosti. Oba klouby pracují jako celek a jakákoliv změna u jednoho kloubu ovlivňuje funkci druhého kloubu. TMK se skládá ze dvou kostních komponent – condyl mandibulae a fossa glenoidalis ze spánkové kosti. Kondyl má elipsovitý tvar s mediolaterálním průměrem 20mm a anteroposteriorním průměrem 10 mm. Podélné osy obou kondylů svírají úhel 150 až 160°. Kloubní povrch je pokryt vazivovou chrupavkou, místo hyalinní chrupavky, která je obvykle přítomna v synoviálním kloubu. Vazivová chrupavka je odolnější proti degenerativním změnám a má lepší regenerační vlastnosti. Důležitou součástí TMK je jeho disk (discus articularis), který má bikonkávní tvar a vyplňuje prostor mezi kondylem dolní čelisti a fossou temporalis. Díky specifickému a komplikovanému složení disku je jeho funkce především jako tlumič veškerých nárazů a otřesů. Funkce kloubu je symetrická a harmonizace pohybů umožňuje kousání, žvýkání a mluvení. Rozlišujeme několik pohybů – deprese, elevace dolní čelisti, propulze a lateropulze [1]. Porucha tempo-

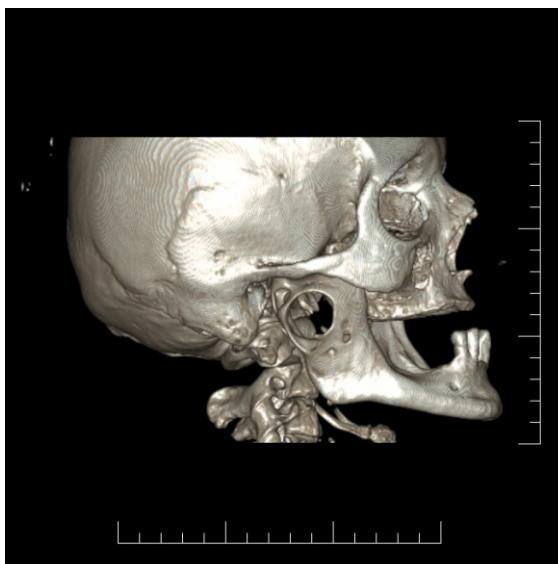
mandibulárního kloubu je pojem zahrnující řadu klinických problémů, které postihují žvýkací svalstvo, čelistní kloub a související struktury. Tyto poruchy jsou doprovázeny symptomy – bolest žvýkacích svalů, bolest TMK při pohybu nebo v klidu a bolest souvisejících tvrdých a měkkých tkání. Mezi další příznaky patří omezené otevírání nebo uchylování dolní čelisti, zvukové fenomény a bolesti hlavy [2]. Incidence poruchy TMK v dospělé populaci je 6-12 % a přibližně u deseti miliónů lidí v USA [3]. U většiny těchto pacientů je dostatečná konzervativní terapie, ale u pacientů, kde postižení TMK dosáhlo pokročilejšího stadia je chirurgická terapie nezbytná. Nejčastější etiologie nejzávažnějších poruch TMK je stav po úrazu, osteoartróza, artritida, ankylóza, idiopatická kloubní resorpce, autoimunitní onemocnění (např. revmatoidní artritida, psoriatická artritida, lupus, sklerodermie, Sjögrenův syndrom, ankylotická spondylitida) [4]. Mezi nejméně zatěžující chirurgické ošetření patří vložení alloplastických implantátů, které mají za úkol zlepšení funkčnosti TMK. Při rozsáhlém poškození TMK a výrazném omezení funkce je nutný chirurgický zákrok, kdy je čelistní kloub (TMK) nahrazen kloubní náhradou. Subtotální náhrada je v případě, kdy nahrazujeme kondyl dolní čelisti nebo fossu temporalis, pokud by byly nahrazeny obě části, jednalo by se již o totální náhradu [5]. V dnešní době moderní pokrok v biomedicině umožňuje lépe pochopit funkci temporomandibulárního kloubu (TMK) a to pomocí zobrazovacích metod. Zpracování získaných dat pomocí moderních programů vytváří nový pohled na rekonstrukci TMK, dále zlepšuje diagnostickou a terapeutickou fázi. K zobrazení TMK jsou k dispozici různé zobrazovací metody.

Mezi ně patří základní rtg snímky (např. Ortopantomogram, tomogramy TMK), ultrasonografie, MRI, spirální a 3D-počítačová tomografie. Na základě těchto metod lze snímat požadované úseky čelistního kloubu v různých rovinách a ze získaných dat vytvořit model TMK ve 3D rekonstrukci. To přináší zkvalitnění diagnostického procesu [6]. Cílem naší studie je ukázka využití zobrazovacích metod při sledování pacientky během 1 roku.

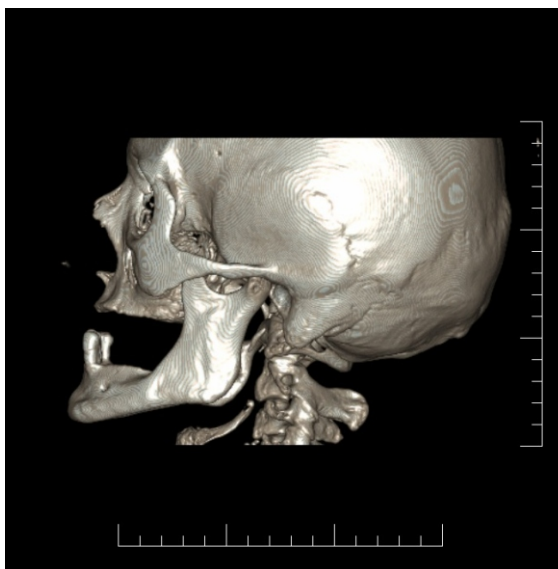
2. Kazuistika

75 letá pacientka byla odeslána na naši kliniku ze stomatologie v Českých Budějovicích pro rozsáhlou susp. radikální cystou dolní čelisti vpravo. Od roku 2002 opakovaně ošetřena v soukromé stomatologické ambulanci pro intraorální píštěl v dané lokalitě. Vždy byly obtíže doprovázeny otokem a kontrakturou, které pozitivně reagovaly na ATB terapii. Po výkonu se objevuje hypostézie pravé poloviny dolního rtu. Do roku 2009 se výše popsané potíže objevily celkem ještě třikrát. Na CT nález byl patrný rozsáhlý cystický útvar destruující větev mandibuly vpravo v rozsahu od kloubního výběžku (vč. svalového) až po úhel mandibuly. Při ambulantním vyšetření patrná asymetrie obličeje vpravo v subzygomatické krajině, kloubní hlavice TMK l.d. nehmátná, otvírání úst volné, dutina ústní bez patologického nálezu. Bezzubý horní oblouk, v dolním oblouku parodontické zuby 33, 43, 44. Chrup sanován snímatelnými protetickými náhradami. Hypostézie n.V3.l.d. Kvůli výrazné destrukci ramus mandibulae a signifikantnímu oslabení kraniokaudálně a mediolaterálně je pacientka ohrožena rizikem patologické fraktury, proto byla navržena léčba subtotální náhradou TMK l.d. (Obr. 1,2,3).

V červnu 2009 v klidné celkové anestézii (NTI) byla provedena exstirpace cysty mandibuly vpravo, parciální resekce mandibuly s přerušením kontinuity v těle mandibuly a exartikulace. Poté provedena rekonstrukce TMK l.d. subtotální náhradou (UNI-LOCK). Odebraný materiál odeslán na histologické vyšetření, jehož výsledek potvrdil diagnózu radikální cysty.



Obr. 1. 3D-CT snímek s rozsáhlou cystou v ramus mandibulae l.d.



Obr. 2. 3D-CT snímek zdravého TMK l.sin.



Obr. 3. CT rozsah cysty před výkonem.

V průběhu hospitalizace pacientka zajištěna ATB celkem na 10 dní (Augmentin), antiedematózní kortikoterapií (Dexamed) a analgetiky (Morphin, Novalgin). 4.den po výkonu pacientka začíná s rehabilitací otvírání úst 10min á 1hod pomocí dřevěných špátlí. 7.den po výkonu propuštěna do domácího ošetřování i nadále pokračuje RHB (Obr. 4). Po propuštění následovaly 3 kontroly v 3-týdenních intervalech, kdy pacientka v dobré kondici, zevně bez otoků, rány zhojeny, otvírání úst se postupně zlepšuje. Dne 23. 2. 2010 provedeno kontrolní CT. Nyní pacientka bez obtíží, bez deformace obličeje, parestézii n.V3l.dx. neguje, otvírání úst na 30mm, čelist při otvírání úst neuchyluje. Pacientka na subtotální náhradu TMK l.dx. psychicky i fyzicky dobře adaptována, toho času zařazena do dispenzarizace. Díky 3D-CT vizualizaci mohla být chirurgická léčba redukována na minimum. S odstupem jednoho roku udává pacientka minimální obtíže (Obr. 5, 6)

3. Metody

Data pacienta byla zpracována (filtrování, ořezávání, spojení, zpracování Iso-surface, registrování) a zobrazena za pomoci softwaru CTVolRegApp, který byl vyvinut na Západočeské univerzitě za podpory Multipod Application Framework (MAF) [7], což je zobrazovací systém založený na VTK [8], ale také na dalších specializovaných databázích. CT snímač bude využit k sledování pacientů s anomálií v orofaciální oblasti. Tyto snímače jsou založeny na objemové tomografii s využitím rozšířeného 2D paprsku, sloužícího k určení lokalizace. Toto je kombinováno s 3D paprskem. Tato technika zahrnuje 360 stupňové snímání, kdy je využita kombinace paprsku a lokátoru. Detektor se synchronizovaně pohybuje kolem pacientovy hlavy, která je pevně fixována. Obrazy, které nazýváme základními, jsou získávány v určitých stupňových intervalech. Tyto snímky jsou podobné snímkům získaným z laterální kefalometrie, které jsou dělány v rychlých sekvencích. Tyto základní obrazy bereme jako referenční. Tato data jsou vložena do programu s integrovanými algoritmy, zahrnující zpětnou projekci, které mohou být využity k poskytnutí rekonstrukce obrazů ve třech na sebe kolmých rovinách.



Obr. 4. Subtotální náhrada TMK.



Obr. 5. 3D-CT zobrazení subtotální protézy 1 rok po zákroku.



Obr. 6. 3D-CT zdravého TMK l.sin. 1 rok po zákroku.

Ačkoliv 3D-CT principy se užívají již po dvě dekády, je to nedávno, kdy díky rozvoji vysoce kvalitních detektorů a výkonných počítačů, jsou tyto systémy finančně dostupné [9].

4. Výsledek

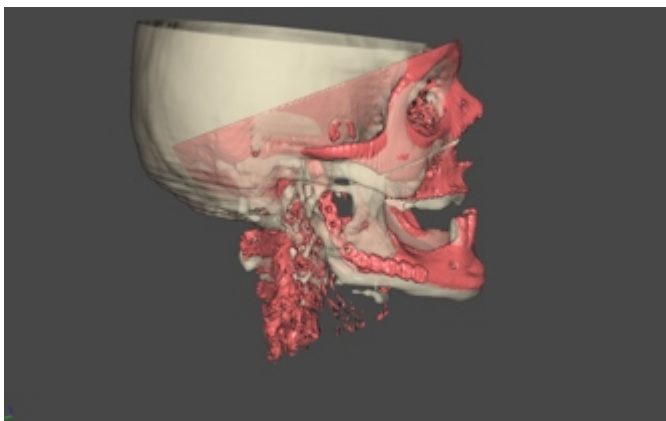
Zpracovali jsme CT snímky TMK před a po léčbě. Snímky byly zhotoveny pomocí různých 3D-CT snímačů s různou kvalitou (Obr. 7, 8). Obr. 7: na tomto 3D-modelu sledujeme postiženou pravou stranu, kde byla přítomna rozsáhlá radikulární cysta. V oblasti větve a kloubního výběžku je viditelná destrukce kostní tkáně, která byla nahrazena subtotální náhradou. Obr. 8: zde je patrný pohled na zdravou stranu dolní čelisti. Obr. 9: zde na modelu pozorujeme subtotální náhradu a fixaci ve zdravé kostní tkáni, kde náhrada přesně anatomicky kopíruje tvar. Poloha hlavy byla u obou zhotovených CT snímků různá. Pro analýzu získaných dat stavu před a po výkonu, jsme museli zpracovat jednotlivá data v softwarovém programu, který vytvořil 3D-CT modely z obou měření. Překrytím obou modelů do jednoho nám umožňuje sledovat změny na dolní čelisti, které nastaly po zavedení subtotální náhrady čelistního kloubu. Oba modely jsou označeny různou barvou, proto lze sledovat změny na společném modelu. S modelem lze prostorově manipulovat (např. rotace, translace).

5. Diskuze

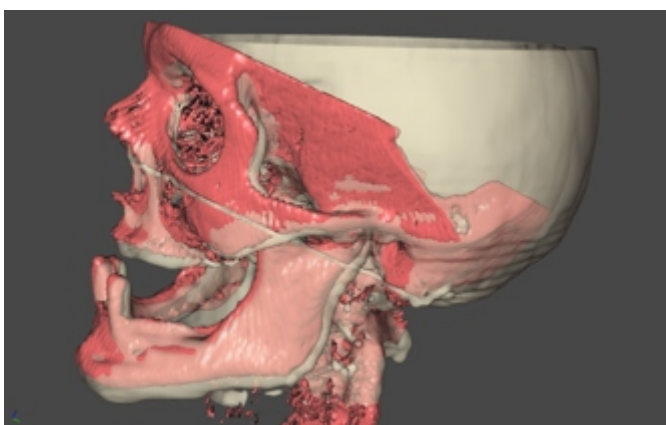
Matematická a počítačová modelace dolní čelisti a čelistního kloubu má potenciál pro využití v mnoha oblastech klinické praxe, v preprotetické chirurgii a v maxilofaciální chirurgii (např. u náhrady TMK). Klinické vyšetření pacientů, u kterých byla provedena jednostranná resekce s následnou rekonstrukcí TMK, nám prokázala, že přítomnost bolesti spojené s náhradou TMK je minimální. Matematické modelování a matematická simulace náhrady čelisti a/nebo TMK může být využito v plánování zhotovení kloubní náhrady. Další aplikace matematického modelování čelisti může být využito k zobrazení změn po ortognátní operaci, k hodnocení vlivu okluzních dlah na zatížení TMK, stejně jako k hodnocení u nádorových stavů a cyst. Naše studie navazuje na tuto publikaci [10].

Vývoj CT techniky, který se zaměřil na zlepšení kvality a usnadnění manipulací se snímky, zdokonaluje zobrazení všech struktur. Standardizace obrazových formátů na základě speciálních protokolů a využívání softwarů vede k častějšímu využití zobrazovacích metod v klinické fázi.

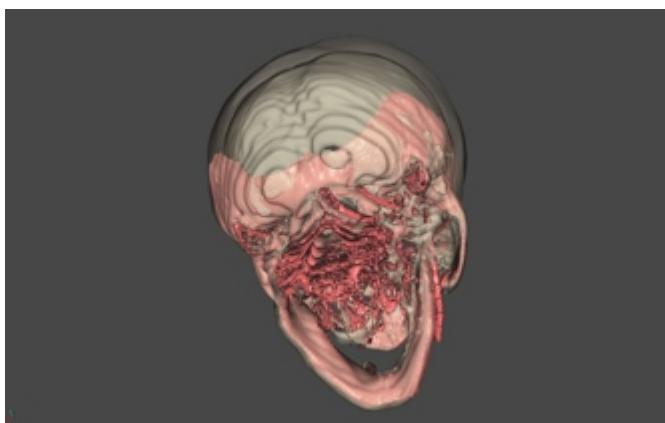
Manipulace s jednotlivými snímky by měla být jednoduchá pro zdravotníky, kteří pracují s diagnostikou, plánováním a léčbou pacientů [11]. Naše studie prokázala praktické využití 3D-CT snímků během léčby TMK.



Obr. 7. 3D-CT zobrazení postižené pravé strany TMK l.dx. před a po léčbě.



Obr. 8. 3D-CT zobrazení zdravého TMK l.sin před a po léčbě.



Obr. 9. 3D-CT model crania před a po léčbě.

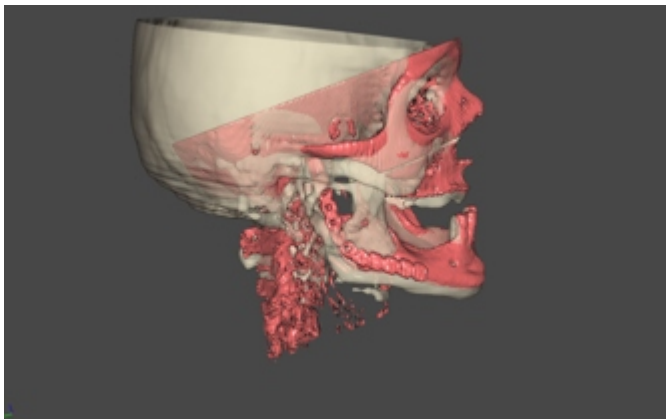
Ačkoliv mnoho poruch čelistního kloubu a souvisejících struktur může být diagnostikováno klinicky na základě anamnézy a fyzikálního vyšetření, jsou onemocnění TMK, která můžeme diagnostikovat jen na základě zobrazovací techniky pro přesnou diagnózu a určení rozsahu postižení. Zobrazení tvrdých a měkkých tkání čelistního kloubu může být použito nejen pro léčbu, plánování, ale také pro hodnocení dlouhodobé léčby.

Poděkování

Tato práce byla zhotovena za podpory grantu 9991-3 IGA MZČR, projektu FI-IM5/142 a SVV-2010-265513.

Literatura

- [1] Hlinakova P., Dostálová T., Danek J., Nedoma J.: „Temporomandibular joint and its two-dimensional and three-dimensional modelling“. Mathematics and computers in simulation Vol.80, pp. 1256- 2010, 2010.
- [2] Fikáčková H., Dostálová T., Navrátil L., Klaschka J.: Effectiveness of low-level laser therapy in temporomandibular joint disorders: A placebo-controlled study. Photomedicine and Laser Surgery, vol. 25, pp. 297-303, 2007.
- [3] Lipton J.A., Ship J.A., Larach-Robinson D.: Estimated prevalence and distribution of reported orofacial pain in the United States. J Am Dent Assoc vol. 124, pp. 11521, 1993.
- [4] Plesh O., Sinisi S.E., Crawford P. B., Gansky S.A.: Diagnoses based on the research diagnostic criteria for temporomandibular disorders in a biracial population of young women. J Orofac Pain. vol. 19, pp. 6575, 2005.
- [5] Wolford L.M., Karras S.C.: Autologous fat transplantation around temporomandibular joint total joint prostheses: preliminary treatment outcomes. J. Oral Maxillofac Surg. vol. 55, pp. 45251, 1997.
- [6] Wadhwa S., Kapila S.: TMJ disorders: future innovations in diagnostics and therapeutics. J Dent Educ. vol. 72, pp. 930947, 2008.
- [7] Viceconti M., Zannoni C., Testi D., Petrone M., Perticoni S., Quadrani P., Taddei F., Imboden S., Clapworthy G.J.: The Multimod Application Framework: a rapid application development tool for computer aided medicine. Comput Methods Programs Biomed, 85: 138-151, 2007
- [8] Schroeder W., Martin K., Lorensen B.: The Visualization Toolkit: an object-oriented approach to 3D graphics. Prentice Hall (1998)
- [9] Scarfe W. C.: Clinical applications of cone-beam computed tomography in dental practice. J Can Dent Assoc, vol. 72(1), pp. 7580, 2006.
- [10] Danek J., Hlinakova P., Preckova P., Dostalova T., Nedoma J.: Three-dimensional modelling of the temporomandibular joints and of their artificial replacements. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 6019 LNCS (PART 4), pp. 62-71, 2010
- [11] Farman A.G.: Raising standards: digital interoperability and DICOM. Oral Surg Oral Med Oral Pathol Oral Radiol Endod. vol. 99, pp. 525-6, 2005.



Obr. 10. CT snímek náhrady 1 rok po léčbě.

Kontakt

MUDr. Eliška Feltlová

Dětská stomatologická klinika 2. LF UK
a FN Motol
V úvalu 84
150 06 Praha 5
e-mail: nelis.elis@seznam.cz

Metody analýzy přežití

J. Fürstová¹

¹Lékařská fakulta, Univerzita Palackého v Olomouci

Školitel: Doc. Zdeněk Valenta, M.Sc., M.S., Ph.D.

Souhrn

Analýza přežití je soubor statistických metod, ve kterých je zkoumána doba do výskytu sledované události. V biomedicínských aplikacích takovou událostí může být např. výskyt primárního onemocnění nebo úmrtí pacienta. Charakteristickým jevem v analýze přežití je tzv. cenzorování a krácení dat, kdy pracujeme pouze s částečnou informací o přežití pacientů (např. sledování pacientů typicky skončí dřív, než všichni zemřou). Metody analýzy přežití slouží k odhadu rozdělení doby do výskytu sledované události, resp. přežití pacienta, k určení rizikových faktorů, které ovlivňují délku přežití, a také k predikci času přežití v závislosti na přítomnosti rizikových faktorů. K rozvoji metod analýzy přežití přispívá také využití teorie čítacích procesů a martingalů. Jednorozměrné metody analýzy přežití nelze použít, jestliže není splněn předpoklad nezávislosti jednotlivých dob přežití. V takovém případě je nutné využít vícerozměrné metody, jako jsou např. vícestavové modely nebo modely náchylnosti.

Klíčová slova: analýza přežití, funkce přežití, riziková funkce, kumulativní riziková funkce, cenzorování, krácení, Kaplanova-Meierova funkce přežití, Nelsonova-Aalenova funkce rizika, Coxův model proporcionálních rizik, metoda parciální věrohodnosti, čítací proces, historie, filtrace, martingal, konkurující rizika, vícestavové modely, modely náchylnosti

1. Úvod

Analýza přežití je souborem statistických metod, které zkoumají dobu do výskytu sledované události, kterou v biomedicině typicky představuje výskyt onemocnění nebo úmrtí pacienta, odtud analýza přežití. Počátky analýzy přežití sahají do osmnáctého století, kdy byly zaznamenány první analýzy lidské úmrtnosti. Během 2. světové války byla tato analýza zaměřena převážně na oblast strojírenství,

prioritou se stala životnost vojenské výzbroje. V technických oborech se nazývá analýzou spolehlivosti (angl. „reliability“). Po 2. světové válce se zájem opět stočil k lidem do zdravotnictví a nově též do ekonomie. Od šedesátých let 20. století, kdy vyšel fundamentální článek E. L. Kaplana a P. Meiera [9], je hlavní aplikací této analýzy oblast medicínského a farmaceutického výzkumu, zejména v oboru klinických studií.

2. Základní pojmy analýzy přežití

2.1 Funkce přežití a riziková funkce

Nechť X značí dobu do výskytu určitého jevu, t.j. X je (absolutně) spojitá nezáporná náhodná veličina s konečnou střední hodnotou. Rozdělení veličiny X lze popsat několika způsoby:

• Hustota pravděpodobnosti $X: f(x), x \geq 0$.

• Funkce přežití:

$$S(x) = P(X > x) = \int_x^\infty f(u) du \\ = 1 - F(x),$$

kde $F(x)$ je distribuční funkce X . Funkce přežití tedy značí pravděpodobnost, že subjekt bude žít déle než do času x (tedy že u něj sledovaná událost nastane později než v čase x).

• Riziková funkce:

$$\lambda(x) = \lim_{\Delta x \rightarrow 0^+} \frac{P(x \leq X < x + \Delta x \mid X \geq x)}{\Delta x},$$

pro všechna $x > 0$. Riziková funkce udává derivaci podmíněné pravděpodobnosti výskytu sledované události v příštím okamžiku za předpokladu, že událost nenastala do času x . Mezi rizikovou funkcí a funkcí přežití je jednoznačný vztah:

$$\lambda(x) = \lim_{\Delta x \rightarrow 0^+} \frac{1}{\Delta x} \frac{P(x \leq X < x + \Delta x)}{P(X \geq x)} \\ = \lim_{\Delta x \rightarrow 0^+} \frac{\frac{1}{\Delta x} \int_x^{x+\Delta x} f(u) du}{\int_x^\infty f(u) du} = \frac{f(x)}{S(x)}$$

$$\frac{dS(x)}{S(x)} = -\frac{dx}{dx} \ln S(x).$$

• Kumulativní riziková funkce:

$$\Lambda(x) = \int_0^x \lambda(u) du = -\ln S(x).$$

Platí tedy

$$S(x) = \exp(-\Lambda(x)) = \exp\left(-\int_0^x \lambda(u) du\right).$$

Je-li X diskrétní náhodná veličina s hodnotami $x_1 < x_2 < \dots$ a rozdělením $f(x_i) = P(X = x_i)$, $i = 1, 2, \dots$, pravděpodobnosti je funkce přežití tvaru

$$S(x) = \sum_{j: x_j > x} f(x_j)$$

a riziková funkce v bodě x_i je

$$\lambda_i = P(X = x_i \mid X \geq x_i) = \frac{f(x_i)}{S(x_i^-)}, i = 1, 2, \dots,$$

kde $S(x^-) = \lim_{t \rightarrow x^-} S(t)$. Diskrétní funkci přežití a rozdělení pravděpodobnosti lze také zapsat formou součinu (viz [8]):

$$S(x) = \prod_{j: x_j \leq x} (1 - \lambda_j), \quad f(x_i) = \lambda_i \prod_{j=1}^{i-1} (1 - \lambda_j).$$

Rozdělení veličiny X může být i obecnější, zčásti diskrétní a zčásti spojitý. K oběma těmito částem lze přistupovat současně, a to skrze součinný integrál: Nechť $\lambda_c(x)$ je spojitá část rizikové funkce a nechť $\lambda_1, \lambda_2, \dots$ jsou diskrétní složky v časech $x_1 < x_2 < \dots$. Celková funkce přežití je potom

$$S(x) = \exp\left(-\int_0^x \lambda_c(u) du\right) \prod_{j: x_j \leq x} (1 - \lambda_j)$$

a kumulativní riziková funkce je tvaru

$$\Lambda(x) = \int_0^x \lambda_c(u) du + \sum_{j: x_j \leq x} \ln(1 - \lambda_j).$$

Nechť $d\Lambda(x)$ značí diferenciální přírůstek kumulativního rizika na intervalu $[x, x + dx)$. Potom

$$\begin{aligned} d\Lambda(x) &= \Lambda(x^- + dx) - \Lambda(x^-) \\ &= P(X \in [x, x + dx) | X \geq x) \\ &= \begin{cases} -\ln(1 - \lambda_i) & \text{for } x = x_i, i = 1, 2, \dots \\ \lambda_c(x)dx & \text{otherwise.} \end{cases} \end{aligned}$$

Funkci přežití v diskretní, spojitě nebo smíšené formě lze potom zapsat jako

$$S(x) = P_0^x(1 - d\Lambda(u)),$$

kde

$$P_0^x(1 - d\Lambda(u)) = \lim_{r \rightarrow \infty} \prod_{k=1}^r (1 - (\Lambda(u_k) - \Lambda(u_{k-1})))$$

je součinový integrál [8].

2.2 Cenzorování a krácení

Analýza přežití se odlišuje od ostatních statistických metod tzv. cenzorováním. K cenzorování dat dochází, pokud máme neúplné informace o přežití pacientů. U některých pacientů totiž nemusíme znát přesnou dobu výskytu události, pouze víme, že událost nastala v nějakém intervalu. Rozlišujeme několik typů cenzorování: cenzorování zprava, zleva a intervalové cenzorování. V biomedicínských aplikacích se nejčastěji vyskytuje cenzorování zprava. To nastává, jestliže k nekompletnosti časového údaje dojde na pravé straně sledovaného intervalu, tedy jestliže je studie ukončena dříve, než u všech pacientů dojde ke sledované události nebo pokud pacient nemůže být déle sledován (zemře z jiného důvodu, odstoupí ze studie, odstěhuje se atd.)

Nechť X_1, X_2, \dots, X_n jsou nezávislé, stejně rozdělené doby přežití a nechť C_1, C_2, \dots, C_n jsou nezávislé, stejně rozdělené doby cenzorování. Doba přežití X_i příslušná i -tému subjektu bude známa, právě když $X_i < C_i$. Jestliže $C_i < X_i$, pak bude doba do výskytu sledované události u i -tého subjektu cenzorována v C_i . Je tedy vhodné reprezentovat informaci o přežití pomocí dvojice náhodných veličin (T_i, δ_i) , kde $T_i = \min(X_i, C_i)$ a $\delta_i = I(X_i < C_i)$ a I je indikátor výskytu události ($\delta_i = 1$ pokud událost nastala, $\delta_i = 0$ v případě cenzorování).

Dalším jevem typickým pro analýzu přežití

je krácení (truncation). Krácení dat se objevuje v případech, kdy jsou do studie zahrnuti pouze ti pacienti, jejichž událost nastala v daném intervalu (T_L, T_R) . Při krácení zleva je $T_R = \infty$, při krácení zprava je $T_L = 0$. O pacientech, jejichž událost nastala mimo interval (T_L, T_R) nemáme žádné informace. Rozdíl mezi cenzorováním a krácením je ten, že v případě cenzorování máme o každém pacientovi alespoň částečné informace. V případě krácení je zapotřebí při sestavování věrohodnostní funkce použít podmíněnou distribuci (viz [10]).

Při sestavování věrohodnostní funkce je zásadním předpokladem nezávislost doby přežití a doby cenzorování. Cenzorování je nezávislé, jestliže v každém čase $t > 0$ je riziková funkce stejná, jako kdyby k žádnému cenzorování nedošlo [8]. Tedy pro každé t je zapotřebí

$$\lim_{\Delta t \rightarrow 0} \frac{P(T \in [t, t + \Delta t) | T \geq t)}{\Delta t} = \lim_{\Delta t \rightarrow 0} \frac{P(T \in [t, t + \Delta t) | T \geq t, Y(t) = 1)}{\Delta t},$$

kde $Y(t) = 1$ značí, že subjekt je v riziku výskytu sledované události v čase t (zatím u něj nenastala událost ani nebyl údaj o době do výskytu události cenzorován před časem t).

2.3 Čítací procesy a martingaly

Čítací procesy představují alternativní přístup k datům o době přežití. Čítací proces $N = \{N(t), t \geq 0\}$ je náhodný proces, pro nějž $N(0) = 0$ a jehož hodnota v čase t udává počet událostí, které nastaly v intervalu $(0, t]$. Trajektorie (realizace) procesu N je neklesající, zprava spojitá schodovitá funkce, která má skok v každém čase, kdy nastane událost (události). Ve formulaci čítacích procesů je dvojice (T_i, δ_i) , zavedená v sekci 2.2, nahrazena dvojicí funkcí $N_i(t), Y_i(t)$, $i = 1, \dots, n$ kde $N_i(t)$ = počet událostí subjektu i zaznamenaných v $(0, t]$

$$Y_i(t) = \begin{cases} 1 & \text{subjekt } i \text{ je v riziku v čase } t, \\ 0 & \text{jinak.} \end{cases}$$

$N_i(t)$ je čítací proces, zatímco $Y_i(t)$ je predikovatelný proces, tedy proces, jehož hodnota v čase t je známa již těsně před t v čase t . Tento proces má zleva spojitě trajektorie. Zprava cenzorovaná data jsou v této formulaci zahrnuta jako zvláštní

případ:

$$N_i(t) = I(T_i \leq t, \delta_i = 1) \text{ a } Y_i(t) = I(T_i \geq t).$$

Historie čítacího procesu (neboli filtrace) v sobě obsahuje veškeré do daného okamžiku známé informace. Přesněji, historie $\{F_t, t \geq 0\}$ je σ -algebrou generovanou procesy N_i a Y_i :

$$F_t = \sigma(N_i(s), Y_i(s^+), i = 1, \dots, n; 0 \leq s \leq t),$$

$$\text{kde } Y_i(s^+) = \lim_{u \rightarrow s^+} Y_i(u).$$

Tedy F_t obsahuje informace do doby t včetně. Informací v F_t přibývá s postupujícím časem, t.j. $F_s \subseteq F_t$ pro $s \leq t$ [4]. Nechť $dN_i(t)$ značí přírůstek N_i v intervalu $[t, t + dt)$:

$$dN_i(t) = N_i((t + dt)^-) - N_i(t^-).$$

Pro každé $t > 0$ nechť

$$F_{t-} = \sigma(N_i(s), Y_i(s), i = 1, \dots, n; 0 \leq s < t)$$

je historie procesů $N_i(s), Y_i(s), i = 1, \dots, n$ do doby t (vyjímaje). Potom platí (viz [4]):

$$E(dN_i(t) | F_{t-}) = Y_i(t) \lambda_i(t) dt,$$

kde $\lambda_i(t)$ je riziková funkce nebo funkce hazardu (též intenzita procesu). Proces

$$\Lambda_i(t) = \int_0^t Y_i(s) \lambda_i(s) ds, \quad t \geq 0,$$

se proto nazývá procesem změn intenzity výskytu sledované události. Pro libovolné pevné t je výsledkem tohoto procesu náhodná veličina, jež aproximuje počet skoků (tj. výskytů sledované události) čítacího procesu N_i v intervalu $(0, t]$ Pro tyto procesy platí $EN_i(t) = E\Lambda_i(t)$ a tedy

$$E(N_i(t) | F_{t-}) = E(\Lambda_i(t) | F_{t-}) = \Lambda_i(t)$$

Uvažujme proces

$$M_i(t) = N_i(t) - \int_0^t Y_i(s) \lambda_i(s) ds, \quad t \geq 0, i = 1, \dots, n. \quad (1)$$

Tento proces lze ekvivalentně zapsat jako

$$M_i(t) = \int_0^t dM_i(s),$$

kde

$$dM_i(t) = dN_i(t) - Y_i(t) \lambda_i(t) dt.$$

Lze dokázat, že $E(dM_i(t)|F_t) = 0$ pro všechna t a $E(M_i(t)|F_s) = M_i(s)$ pro všechna $s \leq t$ [4]. Proces, který splňuje tyto (ekvivalentní) podmínky, se nazývá *martingal*. Podle Doob-Meierovy dekompoziční věty (viz [4]) lze každý čítací proces jednoznačně rozložit na součet martingalu a kompenzátoru C , což je predikovatelný, zprava spojitý proces, pro nějž $C(0) = 0$. Z rovnice (1) dostáváme

$$\begin{aligned} N_i(t) &= M_i(t) + \int_0^t Y_i(s) \lambda_i(s) ds \\ &= M_i(t) + \Lambda_i(t), \end{aligned} \quad (2)$$

kde $M_i(t)$ je martingal příslušný čítacímu procesu $N_i(t)$ a $\Lambda_i(t)$ je kompenzátor procesu N_i vzhledem k filtraci F_t . Proces (2) lze ekvivalentně zapsat formou diferenciálních přírůstků

$$dN_i(t) = dM_i(t) + Y_i(t) \lambda_i(t) dt.$$

Využití čítacích procesů a martingalů zjednodušuje značení. Kromě toho lze vybudovanou teorii martingalů s výhodou použít v analýze přežití zejména při asymptotických metodách odvozování odhadů a testování hypotéz.

3 Neparametrické a semiparametrické modely

Jedním z hlavních úkolů analýzy přežití je odhad základních funkcí (funkce přežití a rizikové funkce). K tomu lze využít parametrické metody, při jejichž využití předpokládáme, že data jsou výběrem ze známého rozdělení. Abychom se takovému předpokladu vyhnuli, je vhodné použít neparametrické modely. Nejjednodušším neparametrickým odhadem distribuční funkce je empirická distribuční funkce

$$F_n(x) = \frac{\text{počet pozorování} \leq x}{n},$$

kdy spojitou distribuční funkci odhadujeme diskrétní distribuční funkcí vypočtenou na základě naměřených dat. Máme-li necenzorovaná data o n pacientech, funkci přežití lze odhadnout jako $S_n(t) = 1 - F_n(t)$. Problém nastane, pokud jsou data cenzorována. Standardní metody totiž cenzorování neumožňují. Rozvoj této oblasti statistiky byl stimulován články Kaplan a Meiera [9] a Coxe [3].

3.1 Kaplanův-Meierův odhad funkce přežití a Nelsonův-Aalenův odhad funkce kumulativního rizika

Kaplanův-Meierův odhad funkce přežití se nejčastěji udává ve tvaru

$$\hat{S}(t) = \prod_{i:t_i \leq t} \left(1 - \frac{d_i}{R_i}\right),$$

kde d_i značí počet událostí zaznamenaných v čase t_i a R_i je počet subjektů v riziku v čase t_i (necenzorovaní pacienti, kteří jsou v riziku těsně před časem t_i). Rozptyl tohoto odhadu odhadujeme pomocí Greenwoodova vzorce (viz [11]):

$$\hat{V}(\hat{S}(t)) = \hat{S}(t)^2 \sum_{i:t_i \leq t} \frac{d_i}{R_i(R_i - d_i)}.$$

Z Kaplanova-Meierova odhadu lze také vyvodit odhad kumulativního rizika:

$$\hat{\Lambda}(t) = -\ln(\hat{S}(t)).$$

Alternativní odhad funkce kumulativního rizika navrhl Nelson v roce 1972 [12]. Aalen tento odhad zavedl znovu v roce 1978 [1] v rámci teorie čítacích procesů. Odhad se proto nazývá Nelsonův-Aalenův odhad kumulativního rizika:

$$\tilde{\Lambda}(t) = \sum_{i:t_i \leq t} \frac{d_i}{R_i}.$$

Rozptyl Nelsonova-Aalenova odhadu odvodil Aalen právě pomocí čítacích procesů:

$$\tilde{V}(\tilde{\Lambda}(t)) = \sum_{i:t_i \leq t} \frac{d_i}{R_i^2}.$$

Z Nelsonova-Aalenova odhadu kumulativního rizika dostáváme též alternativní odhad funkce přežití:

$$\tilde{S}(t) = \exp(-\tilde{\Lambda}(t)).$$

Uvažujme nyní výběr n subjektů z homogenní populace, které sledujeme od času 0. Nechť N_i je čítací proces a Y_i proces, který u i -tého pacienta zaznamenává stav jeho rizika (at-risk process). Nechť

$$N_i(t) = \sum_{s=0}^t N_i(s) \quad \text{a} \quad Y_i(t) = \sum_{s=0}^t Y_i(s), \quad 0 \leq t < \infty.$$

$N_i(t)$ udává celkový počet pozorovaných událostí, zatímco $Y_i(t)$ představuje počet subjektů, které jsou stále v riziku v čase Nelsonův-Aalenův odhad funkce kumulativního rizika má s použitím značení čítacích procesů tvar:

$$\tilde{\Lambda}(t) = \int_0^t \frac{J(u)}{Y(u)} dN_i(u),$$

kde $J(u) = I(Y_i(u) > 0)$ a $0/0$ definujeme jako 0 [8]. Kaplanův-Meierův odhad funkce přežití má potom tvar:

$$\hat{S}(t) = \prod_{u \leq t} (1 - d\tilde{\Lambda}(u)).$$

V analýze přežití často potřebujeme porovnat dvě nebo více skupin pacientů (chceme testovat, zda je významný rozdíl v přežití různých skupin). Za tímto účelem bylo zobecněno několik standardních neparametrických testů tak, aby je bylo možno použít pro cenzorovaná nebo krácená data. Nejznámějšími a nejčastěji používanými jsou *log-rank test*, *Gehanův-Wilcoxonův test* a *Petův-Petův test*. Podrobnější informace o těchto testech lze nalézt např. v [10].

3.2 Coxův model proporcionálních rizik

Obvyklým úkolem klinických studií bývá zjištění vztahu mezi dobou přežití a možnými vysvětlujícími proměnnými. Faktory, které mohou ovlivňovat přežití pacientů, jsou nejčastěji věk, pohlaví, krevní obraz, rodinná anamnéza, stadium onemocnění, atd. Coxův model proporcionálních rizik se stal jednou z nejužívanějších metod pro popis vztahu mezi přežitím pacienta a vysvětlujícími proměnnými. V tomto regresním modelu je proces intenzity (funkce rizika resp. hazardu) pro i -tého pacienta tvaru:

$$\lambda_i(t) = Y_i(t) \lambda_0(t) \exp(\beta^T X_i(t)),$$

kde $Y_i(t)$ je proces stavu rizika, λ_0 je základní riziková funkce (společná pro všechny subjekty ve studii), X_i je vektor vysvětlujících proměnných i -tého pacienta a β je vektor neznámých regresních koeficientů. Coxův model je specifický svým předpokladem, že poměr rizikových funkcí dvou subjektů je konstantní (základní rizikové funkce $\lambda_0(t)$ se v čitateli a jmenovateli vyruší), a tedy vliv vysvětlujících faktorů je zcela očištěn od vlivu času. Regresní koeficienty odhadujeme metodou parciální věrohodnosti (partial likelihood), kterou navrhl Cox v roce 1972 [3].

Parciální věrohodnostní funkce pro parametr β je tvaru:

$$L(\beta) = \prod_{j=1}^k \frac{\exp(\beta^T X_j)}{\sum_{l \in R(t_j)} \exp(\beta^T X_l)},$$

kde $t_1 < \dots < t_k$ jsou necenzorované doby výskytu sledované události, $R(t_j)$ je množina subjektů, kteří jsou v riziku výskytu sledované události v čase t_j (pacienti, u kterých se do doby t_j sledovaná událost nevyskytla ani nebyli cenzorováni) a X_j je vektor vysvětlujících proměnných příslušný pacientovi, u kterého došlo k výskytu události v čase t_j . Hovoříme o metodě parciální věrohodnosti, protože cenzorované subjekty nejsou do věrohodnostní funkce zahrnuty přímo, ale jen prostřednictvím rizikových množin. S parciální věrohodnostní funkcí se nakládá stejně jako s běžnou věrohodnostní funkcí a regresní koeficienty jsou odhadovány standardními prostředky.

4 Vícerozměrné metody analýzy přežití

Jednorozměrné metody analýzy přežití předpokládají, že doby přežití jsou nezávislé a stejně rozdělené. V mnoha aplikacích se však setkáváme s daty, u kterých závislost nelze vyloučit, např. přežití dvojčat či dalších rodinných příslušníků, doba do onemocnění párových orgánů, opakující se událost (např. infarkt nebo infekce), případně více-stavové události (multi-state events). Metody vícerozměrné analýzy přežití tedy pracují s daty, u kterých nelze předpokládat nezávislost jednotlivých dob přežití. Podle monografie [7] se vícerozměrné metody dělí do čtyř hlavních kategorií: více-stavové modely, modely zohledňující náchylnost k výskytu onemocnění (frailty models), marginální modely a neparametrické metody. Data svou strukturou spadají do dvou kategorií: paralelní data, u kterých je počet událostí předem stanoven již od počátku studie, a data longitudinální, ve kterých každému pacientovi odpovídá náhodný počet událostí. Kromě toho rozlišujeme šest typů závislých událostí: více subjektů, podobné orgány, opakované výskyty události, několikanásobná měření, výskyty rozdílných událostí a konkurující rizika (competing risks). Vztah těchto typů

Tab. 1. Přehled datových typů a metod vícerozměrné analýzy; x značí relevantnost, prázdné políčko irrelevantnost. Převzato z [7].

Typ dat	Modely	
	více-stavové	náchylnosti
Více subjektů	x	x
Podobné orgány	x	x
Opakované výskyty události	x	x
Několikanásobná měření		x
Výskyty rozdílných událostí	x	
Konkurující rizika	x	

datových sad ke dvěma hlavním metodám vícerozměrné analýzy (více-stavovým událostem a modelům náchylnosti) je uveden v Tabulce 1. V tomto článku jsou představeny pouze dva výše uvedené přístupy k vícerozměrným datům. Bližší informace o marginálních modelech a neparametrických metodách lze nalézt např. v monografiích [7], [16], nebo [8].

4.1 Konkurující rizika a více-stavové modely

K analýze longitudinálních dat se nejčastěji používají více-stavové modely. Jsou založeny na náhodných procesech, které se v každém čase nacházejí v jednom z daných stavů. V biomedicínských aplikacích jsou těmito stavy např. stav „zdravý“, „nemocný“, „zemřelý“. Mezi jednotlivými stavy se dá přecházet buď obousměrně (zdravý – nemocný) nebo jednosměrně (zdravý – zemřelý, nemocný – zemřelý, zemřelý je tedy tzv. terminální stav). Příkladem více-stavového modelu jsou také konkurující rizika (competing risks). V případě konkurujících rizik sledujeme několik rozdílných příčin úmrtí, jak „usilují“ pacienti o život. U každého subjektu může „zvíťžít“ jen jedna příčina úmrtí (pacient nemůže zemřít několikrát za sebou). Jakmile tedy nastane jedna událost, jsou vyloučeny všechny ostatní události. Analýza konkurujících rizik se věnuje třem oblastem [8]:

1. Prozkoumat vztah mezi vektorem vysvětlujících proměnných a rizikem jednotlivých typů úmrtí.
2. Zjistit, zda pacienti s vysokým rizikem jedné příčiny úmrtí mají také zvýšené riziko jiného typu úmrtí.
3. Odhadnout riziko jedné příčiny úmrtí při odstranění ostatních příčin úmrtí.

Předpokládejme, že u sledovaného jedince může nastat jeden z m různých typů úmrtí. U každého pacienta známe jeho dobu přežití T a vektor vysvětlujících proměnných X . Celková riziková funkce v čase t je dána

$$\lambda(t, X) = \lim_{\Delta t \rightarrow 0^+} \frac{P(t \leq T < t + \Delta t \mid T \geq t, X)}{\Delta t}.$$

Při analýze konkurujících rizik zavádíme rizikové funkce pro jednotlivé příčiny úmrtí:

$$\lambda_j(t, X) = \lim_{\Delta t \rightarrow 0^+} \frac{P(t \leq T < t + \Delta t, J = j \mid T \geq t, X)}{\Delta t}$$

pro $j = 1, \dots, m$ je náhodná veličina udávající typ úmrtí a $t > 0$ Protože pro každý subjekt může nastat pouze jeden z m typů události, platí

$$\lambda(t, X) = \sum_{j=1}^m \lambda_j(t, X)$$

Pro každý typ úmrtí zvlášť lze spočítat Kaplanovu-Meierovu křivku. Takový odhad ovšem nelze interpretovat jako funkci přežití, proto se tento postup nedoporučuje [7]. Namísto toho se používá tzv. Aalenův-Johansonův odhad, který zobecňuje Kaplanův-Meierův a Nelsonův-Aalenův odhad. Tento odhad v sobě zohledňuje všech m různých typů události (podrobnosti viz [8]).

Coxův model příslušný rizikové funkci pro danou příčinu úmrtí je potom tvaru

$$\lambda_j(t, X) = \lambda_{0j}(t) \exp(\beta_j^T X), \quad j = 1, \dots, m.$$

Základní rizikové funkce λ_{0j} i regresní koeficienty β_j mohou být různé pro různé typy události. Koeficienty β_j pro jednotlivé příčiny úmrtí jsou odhadnuty a porovnány věrohodnostními metodami.

K vícestavovým modelům se tradičně přistupuje pomocí Markovských modelů. Uvažujeme nejdříve homogenní populaci bez vysvětlujících proměnných. Označme $A(t)$ stav, jež nastal v čase $t \geq 0$, kde pravděpodobnostní model $A(t)$ je Markovský proces. Sledované subjekty se pohybují mezi $m > 1$ různými stavy. Je-li náhodně vybraný subjekt ve stavu i v čase t je riziko (neboli intenzita) přechodu z i do j v čase dáno vztahem

$$d\Lambda_{ij}(t) = P[A(t^- + dt) = j | A(u), 0 \leq u < t, A(t^-) = i] \\ = P[A(t^- + dt) = j | A(t^-) = i], \quad t > 0,$$

což platí pro všechna $A(u)$, $0 \leq u < t$ s $A(t) = i$ a $i, j \in \{1, \dots, m\}$, $j \neq i$. Tento proces si nepamatuje předchozí historii, t.j. pro určení rizika přechodu je relevantní pouze aktuální stav, v němž se subjekt nachází [8]. Ve spojitém případě platí $d\lambda_{ij}(t) = \lambda_{ij}(t)dt$ pro všechna $i, j = 1, \dots, m$ je tedy riziková funkce přechodu ze stavu i do stavu j . Odhad kumulativní intenzity $\lambda_{ij}(t)$ je dle [8] následující: Uvažujeme soubor n pacientů, jejichž doba přežití může být zprava cenzorována. Nechť $N_{ijk}(t)$ je zprava spojitý proces, jenž udává počet pozorovaných přechodů ze stavu i do stavu j pro subjekt $k = 1, \dots, n$, $i, j = 1, \dots, m$, $i \neq j$. Nechť $Y_{ijk}(t)$ je příslušný proces stavu rizika. Definujeme filtraci

$$F_t = (N_{ijk}(t), Y_{ijk}(u^+), 0 \leq u \leq t),$$

pro $k = 1, \dots, n$, $i, j = 1, \dots, m$, a předpokládejme, že cenzorování je nezávislé na výskytu sledované události. Platí tedy pro všechna i, j, k a $t > 0$

$$P(dN_{ijk}(t) = 1 | F_{t-}) = Y_{ijk}(t) d\Lambda_{ij}(t),$$

Nelsonův-Aalenův odhad kumulativního rizika $\lambda_{ij}(t)$ je potom dán vztahem

$$d\hat{\Lambda}_{ij}(t) = \frac{dN_{ij}(t)}{Y_{ij}(t)}$$

pro všechna $i \neq j$.

Je-li v modelu přítomen též vektor vysvětlujících proměnných X , je riziková funkce tvaru

$$\lambda_{ijk}(t) = \lim_{dt \rightarrow 0^+} \frac{P(A_k(t^- + dt) = j | A_k(t^-) = i, X)}{dt}.$$

Odvození parametrických i neparametrických metod je analogické k již uvedenému postupu. Podrobnosti lze nalézt v monografii [8].

4.2 Modely náchylnosti

Modely náchylnosti k výskytu sledovaného onemocnění (frailty models) rozšiřují a zobecňují Coxův model proporcionálních rizik. Skrze koncept náchylnosti zavádíme do modelu náhodný jev, který reprezentuje skrytou závislost v datech. Tato závislost vzniká například tím, že analyzujeme rodinné příslušníky (subjekty s podobnou genetickou výbavou), podobné (nebo párové) orgány, nebo na subjektech provádíme opakovaná měření. Závislost v datech může být zjevná, většinou jde ale o závislost skrytou, nepozorovanou. Náchylnost je tedy nepozorovaný náhodný faktor, jenž modifikuje rizikovou funkci jedince nebo skupiny subjektů. Klíčovou myšlenkou těchto modelů je, že „nejnáchylnější“ subjekty umírají dříve než ostatní [16].

Uvažujeme nejdříve dvourozměrný model. Nechť

$$S_{12}(t_1, t_2) = P(T_1 \geq t_1, T_2 \geq t_2)$$

je sdružená křivka přežití příslušná dvěma dobám přežití T_1 a T_2 kde $S_{12}(t, t)$ značí pravděpodobnost, že se oba subjekty dožijí alespoň času t . Marginální křivky přežití jsou potom tvaru:

$$S_1(t_1) = P(T_1 \geq t_1) = S_{12}(t_1, 0)$$

$$S_2(t_2) = P(T_2 \geq t_2) = S_{12}(0, t_2).$$

Jsou-li T_1 a T_2 nezávislé, platí $S_{12}(t_1, t_2) = S_1(t_1)S_2(t_2)$. Sdružená riziková funkce je tvaru

$$\lambda_{12}(t_1, t_2) = \lim_{\Delta t \rightarrow 0^+} \frac{P(T_1 \in [t_1, t_1 + \Delta t), T_2 \in [t_2, t_2 + \Delta t) | T_1 \geq t_1)}{\Delta t^2},$$

zatímco příslušné marginální rizikové funkce mají tvar

$$\lambda_i(t_i) = \lim_{\Delta t \rightarrow 0^+} \frac{P(T_i \in [t_i, t_i + \Delta t) | T_i \geq t_i)}{\Delta t},$$

pro $i = 1, 2$. Abychom mohli do modelu přidat náchylnost, předpokládáme, že doby přežití jsou podmíněně nezávislé. Nebo-li předpokládáme, že T_1 a T_2 jsou nezávislé, až na náhodný faktor Z (náchylnost):

$$S_{12}(t_1, t_2 | Z) = S_1(t_1 | Z)S_2(t_2 | Z).$$

Obvykle se také předpokládá, že Z ovlivňuje rizikovou funkci multiplikativně, tedy že platí:

$$\lambda_i(t_i) = Z\lambda_{0i}(t_i) \quad \text{a} \quad S_i(t_i | Z) = S_{0i}(t_i)^Z$$

pro nějakou základní rizikovou funkci $\lambda_{0i}(t)$ a základní funkci přežití $S_{0i}(t)$. V případě, že do modelu zahrnujeme i známé vysvětlující proměnné X , pro rizikovou funkci použijeme Coxův regresní model

$$\lambda_{0i}(t_i) = \lambda_{0i}(t_i) \exp(\beta^T X_i).$$

Za výše uvedeného předpokladu multiplikativní náchylnosti jsou kumulativní rizika tvaru $\Lambda_i(t_i) = Z\Lambda_{0i}(t_i)$. Pro podmíněnou sdruženou funkci přežití pak platí

$$S_{12}(t_1, t_2 | Z) = S_{01}(t_1)^Z S_{02}(t_2)^Z \\ = \exp(-Z\Lambda_{01}(t_1)) \exp(-Z\Lambda_{02}(t_2)) \\ = \exp(-Z(\Lambda_{01}(t_1) + \Lambda_{02}(t_2))).$$

Jelikož je náchylnost Z skrytý faktor, je potřeba ji z funkce přežití vyintegrovat. K tomu účelu dobře poslouží Laplaceova transformace. Pro náhodnou veličinu Z s hustotou pravděpodobnosti $g(z)$ definujeme transformaci

$$L_g(s) = \int \exp(-sz)g(z)dz = E(\exp(-sZ)).$$

Pro dvourozměrnou funkci přežití tedy dostáváme

$$S_{12}(t_1, t_2) = \int_0^\infty S_{12}(t_1, t_2 | Z)g(z)dz \\ = \int_0^\infty \exp(-Z(\Lambda_{01}(t_1) + \Lambda_{02}(t_2)))g(z)dz \\ = L_g(\Lambda_{01}(t_1) + \Lambda_{02}(t_2)),$$

kde Laplaceovu transformaci vyhodnocujeme v bodě

$$s = \Lambda_{01}(t_1) + \Lambda_{02}(t_2).$$

V aplikacích je běžné předpokládat, že Z má rozdělení s explicitní Laplaceovou transformací. Nejčastěji užívané rozdělení zranitelnosti je gamma rozdělení.

Náhodná veličina Z má rozdělení gamma s parametry k a Θ ($Z \sim \Gamma(k, \Theta)$), jestliže její hustota má tvar

$$g(z) = \frac{\theta^k z^{k-1} \exp(-\theta z)}{\Gamma(k)}, \quad k, \theta > 0 \text{ and } z > 0,$$

se střední hodnotou $EZ = \frac{k}{\theta}$,

a rozptylem $\text{var}Z = \frac{k}{\theta^2}$.

Gamma funkce ve jmenovateli hustoty $g(z)$ je definována jako

$$\Gamma(k) = \int_0^\infty u^{k-1} \exp(-u) du, \quad \text{for } k > 0.$$

a platí pro ni $\Gamma(k+1) = k\Gamma(k)$. Gamma rozdělení velmi dobře popisuje data o přežití a je vhodné také z početního a analytického hlediska [19].

Předpokládejme tedy, že náchylnost Z má gamma rozdělení s parametry $k = \Theta = 1/\sigma^2$. Laplaceova transformace je potom tvaru

$$L(s) = \left(\frac{\theta}{\theta + s} \right)^k,$$

což vede k odhadu funkce přežití (viz [7]):

$$S_{12}(t_1, t_2) = \left(\frac{1}{1 + \sigma^2 \Lambda_{01}(t_1) + \sigma^2 \Lambda_{02}(t_2)} \right)^{1/\sigma^2}.$$

Dvourozměrný model lze rozšířit na vícerozměrný model. Za tím účelem uvažujeme data ve formátu: t_{ij} je pozorovaný čas události u j -tého subjektu v i -té skupině (klastru), X_{ij} jsou příslušné vektory vysvětlujících proměnných. Opět budeme předpokládat, že m_i dob přežití v i -tém klastru je nezávislých, až na náhodný faktor Z_i , a známé vysvětlující proměnné X_{ij} . Sdružené rozdělení dob přežití je tedy součinem marginálních rozdělení. Marginální rizikové funkce potom splňují:

$$\lambda_{ij}(t_{ij} | X_{ij}, Z_i) = Z_i \lambda_{0ij}(t_{ij} | X_{ij}).$$

Použijeme-li pro rizikové funkce Coxův regresní model,

$$\lambda_{0ij}(t_{ij} | X_{ij}) = \lambda_0(t_{ij}) \exp(\beta^T X_{ij}).$$

Mají-li skupinové faktory náchylnosti Z_i , gamma rozdělení, má funkce přežití pro dob přežití ve skupině i tvar:

$$S_i(t_i, X_i) = \int_0^\infty \prod_j S_{ij}(t_{ij} | X_{ij}, Z_i) g(z_i) dz_i,$$

kde $t_i = (t_{i1}, t_{i2}, \dots, K, t_{im_i})^T$, $X_i = (X_{ij})_{m_i \times n}$.

Požitím Laplaceovy transformace získáme vztah (viz [14])

$$S(t_i, X_i) = \left(\frac{1}{\Psi} \right)^{1/\sigma^2},$$

kde

$$\Psi = 1 + \sigma^2 \Lambda_0(t_{i1}) \exp(\beta^T X_{i1}) + \dots + \sigma^2 \Lambda_0(t_{im_i}) \exp(\beta^T X_{im_i}).$$

Náchylnost Z můžeme modelovat i pomocí jiných rozdělení než gamma rozdělení. Více informací o použití a vhodnosti jiných rozdělení lze najít v monografii [7]. Ke zranitelnosti Z také můžeme přistupovat neparametricky. Mít zcela neparametrický odhad funkce přežití je sice lákavé, takové odhady jsou ale matematicky velmi složité a v praxi se neužívají [7].

Statistické modely založené na čítacích procesech vhodné pro modely náchylnosti se liší od těch, které jsme dosud použili. V dřívějších modelech jsme užívali proces intenzity $\lambda(t)$ tvaru

$$\lambda(t) dt = P(dN(t) = 1 | N(s), 0 \leq s < t, X).$$

V tomto vztahu předpokládáme, že čítací proces N má skoky pouze o velikosti jedna. V rekurentních a korelovaných datech ovšem mohou nastat skoky o velikosti větší než jedna (v daném čase můžeme u jednoho subjektu zaznamenat více než jednu událost). Je tedy přirozené uvažovat „průměrný skok“ procesu N v rámci kumulativní intenzity:

$$d\Lambda(t) = E(dN(t) | N(s), 0 \leq s < t, X).$$

Pro tento proces intenzity je potom Coxův model tvaru

$$d\Lambda(t) = d\Lambda_0(t) \exp(\beta^T X).$$

Podrobnosti viz [8].

5. Závěr

Analýza přežití je souborem specifických statistických metod. V tomto článku byl představen krátký přehled těchto metod. Standardně používané jednorozměrné modely byly rozšířeny na vícerozměrné, které jsou vhodné pro paralelní a longitudinální data. Byly představeny dva hlavní koncepty vícerozměrné analýzy přežití: vícecestavové modely a modely náchylnosti.

Práce byla podpořena projektem SVV-2010-265513.

Literatura

- [1] Aalen O. O.: Nonparametric Inference for a Family of Counting Processes. *Annals of Statistics* 6 (1978), 701726.
- [2] Andersen P. K., Gill R. D.: Cox's regression model for counting processes: a large sample study. *The Annals of Statistics* 10 (1982), 11001120.
- [3] Cox D. R.: Regression Models and Life-Tables. *Journal of the Royal Statistical Society B* 34 (1972), 187220.
- [4] Fleming T. R., Harrington D. P.: Counting Processes and Survival Analysis. John Wiley & Sons, New York, 1991.
- [5] Füstová J.: Multivariate Methods of Survival Analysis. Doktorandský den 2010, Matfyzpress, Praha, 2010.
- [6] Gill R. D.: Understanding Cox's regression model: a martingale approach. *Journal of the American Statistical Association* 79 (1984), 441447.
- [7] Hougaard P.: Analysis of Multivariate Survival Data. Springer, New York, 2000.
- [8] Kalbfleisch J. D., Prentice R. L.: The Statistical Analysis of Failure Time Data. John Wiley & Sons, New York, 2002.
- [9] Kaplan E. L., Meier P.: Nonparametric Estimation from Incomplete Observations. *Journal of the American Statistical Association* 53 (1958), 457481.
- [10] Klein J. P., Moeschberger M. L.: Survival Analysis. Techniques for Censored and Truncated Data. Springer, New York, 2003.
- [11] Miller R. G., Gong G., Muñoz A.: Survival Analysis. John Wiley & Sons, New York, 1998.
- [12] Nelson W.: Theory and Applications of Hazard Plotting for Censored Failure Data. *Technometrics* 14 (1972), 945965.
- [13] Prentice R. L., Williams B. J., Peterson A. V.: On the regression analysis of multivariate failure time data. *Biometrika* 68 (1981), 373379.
- [14] Rodríguez G.: Multivariate Survival Models. available at <http://data.princeton.edu/>, cited on April 10, 2010.

- [15] Self S. G., Prentice R. L.: Commentary on Andersen and Gill's "Cox's regression model for counting processes: a large sample study". The Annals of Statistics 10 (1982), 11211124.
- [16] Therneau T. M., Grambsch P. M.: Modeling Survival Data. Extending the Cox Model. Springer, New York, 2000.
- [17] Vaupel J. W., Manton K. G., Stallard E.: The impact of heterogeneity in individual frailty on the dynamics of mortality. Demography 16 (1979), 439454.
- [18] Wei L. J., Lin D. Y., Weissfeld L.: Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. Journal of the American Statistical Association 84 (1989), 10651073.
- [19] Wienke A.: Frailty Models in Survival Analysis. Habilitation. Martin-Luther-Universität Halle-Wittenberg, 2007. available at <http://sundoc.bibliothek.uni-halle.de/habil-online/>

Kontakt**Mgr. Jana Füstová**

Lékařská fakulta

Univerzita Palackého v Olomouci

Tř. Svobody 8

771 26 Olomouc

E-mail: jana.furstova@email.cz

Elektronická zdravotnická dokumentace pro onemocnění temporomandibulárního kloubu – Podpora v rozhodování o terapeutickém procesu

R. Hippmann^{1,2}, M. Nagy^{2,3}, T. Dostálová^{1,2}, J. Zvárová^{2,3}, M. Seydlová^{1,2} a E. Feltlová¹

¹Klinika dětské stomatologie, 2. lékařská fakulta Univerzity Karlovy, Praha

²Centrum biomedicínské informatiky, Praha

³Ústav informatiky AV ČR, v.v.i., Praha

Školitelka: Prof. MUDr. Tatjana Dostálová, DrSc., MBA

Souhrn

Naše experimentální práce se věnuje tvorbě nového typu elektronické zdravotnické dokumentace (EHR) pro oblast stomatologie, obzvláště pro poruchy temporomandibulárního kloubu (TMD) a hlavním cílem je vytvoření komplexního systému, který by byl dobrým základem pro podporu rozhodování v terapii TMD.

Náš EHR systém je založen na Stomatologické znalostní bázi, která je vytvořená v aplikaci MUDR (Multimedia distribuuující EHR) a potom také v MUDR Lite aplikaci. Architektura MUDR Lite je postavena na dvou vrstvách (relační databáze (např. MS SQL server) a uživatelské vrstvě (MUDRLite UI), která je popsána v XML konfigurační složce). Tyto aplikace tvoří reálnou grafickou uživatelskou aplikaci DentCross, která je propojena s ASR (Automatic speech recognition) systémem a TTS (Text-to-speech) modulem. Speciální TMD EHR je částí DentCross a je vytvořena ve stejných aplikacích. Celá tato část využívá klasifikaci RCD/TMD (Research diagnostic criteria for TMD).

Integrovaný Dental-TMJ EHR systém umožňuje záznam komplexních zdravotnických dat pacienta ve stomatologii. Postihuje celý ortognátní systém s jeho všemi složkami (především TMD problematika) v grafické podobě. Záznam a manipulaci s daty zjednodušuje ASR systém a TTS modul. Tato aplikace byla navržena s důrazem na jednoduchost ovládání a schopnost zaznamenat úplnou informaci a měla by pomoci při rozhodování o terapii TMD.

Náš Integrated Dental TMJ systém reprezentuje celou stomatologickou informaci pacienta, která je podstatná pro tvorbu léčebného plánu. Spojení s ASR systémem a TTS modulem zjednodušuje

sběr těchto informací. Informace jsou také přehledně uspořádané v grafické formě. Dalším krokem bude pokračování v klinickém testování celého systému a vylepšení ASR systému a TTS modulu. Nevýhodou systému může být nutná zkušenost s ovládáním této aplikace a fungování ASR systému ve velmi hlučném prostředí.

Klíčová slova: elektronická zdravotnická dokumentace, automatické rozpoznávání řeči, zubní kříž, temporomandibulární kloub, poruchy temporomandibulárního kloubu, strukturovaný záznam dat, stomatologie, datový model, text-to-speech modul, Research Diagnostic Criteria for TMD

1. Úvod

Ve všech oblastech medicíny narůstá objem různých druhů informací. Je tedy velice podstatné mít komplexní informaci o zdravotním stavu pacienta (historie, vyšetření a terapeutická data). Také je velice důležité, aby byla tato informace vždy dostupná a uložená ve vhodné plnohodnotné formě. Všechny tyto požadavky mohou splňovat pouze moderní informatické systémy, obzvláště systémy elektronické zdravotnické dokumentace (EHR). Z těchto důvodů stoupá potřeba po osvojení vhodného EHR systému (EHR-S). Ačkoliv mnohé studie o zahrnutí informačních technologií do pracovního procesu odhalily rozporná data týkající se časové efektivity na tento proces, je pro lékaře hodnocení potencionálního vlivu EHR-S na časovou náročnost práce velice důležité [1]. Další podstatný úkol pro zavedení EHR-S systémů do praxe je jejich orientace na komerční distribuci dat v různých systémech, které by měli být sjednoceny se zaměřením na individuální

potřeby [2]. Obecně by měly EHR-S systémy podporovat plynulou, efektivní a vysoce kvalitně integrovanou zdravotnickou péči pomocí užití strukturovaných vstupních dat [3], [4], interoperability [5] a standartů [6]. Na začátku vývoje různých EHR-S systémů byl publikován Mezinárodní standard ISO/DTR 20514:2005, který definuje EHR jako zdroj informací týkajících se péče o zdravotní stav subjektu v počítačem zpracovatelné formě, ukládané a bezpečně sdílené a přístupné pro více autorizovaných uživatelů. Stejný proces zapojení výpočetní techniky byl započat také ve stomatologii. Nabízí se mnoho grafických přístupů, ale ani jeden z nich není podpořen systémem pro automatické rozpoznávání řeči. Užití hlasových příkazů bylo experimentálně zkoušeno již v devadesátých letech [7] jako vhodná náhrada kontroly pouze pomocí počítače a jeho klávesnicového ovládání. Potřebnost použití kontroly počítače a jiných přístrojů lidským hlasem stoupá hlavně v odvětvích, kde jsou zaměstnány obě ruce, jako je například chirurgie nebo stomatologie [8].

Další výzvou je opravdová komplexnost systému. Toto vede k vytváření doplňkových EHR-S, které jsou věnovány např. tkáním temporomandibulárního kloubu a následně pokrývajícím celý ortognátní systém.

TMD je souhrnný pojem pro symptomy jako jsou temporomandibulární dysfunkce, kraniomandibulární onemocnění a mandibulární dysfunkce.

TMD zahrnují celou paletu stavů spojených s bolestivostí a dysfunkcemi temporomandibulárního kloubu a žvýkacích svalů [9]. Předpokládaný výskyt v populaci je asi 20 %. Z toho je asi jen 10% až 20% pacientů, kteří vyhledají terapii [10], [11]. Mezi hlavní projevy patří bolestivost žvýkacích svalů, limitace nebo deviace v otvírání úst a zvuky při pohybu kloubu. Etiologie TMD je multifaktoriální a proto je pro indikování vhodné terapie velice důležitá detailní a komplexní vstupní informace. Diagnostika TMD je obvykle založená na přítomných znacích a symptomech. Pro klasifikaci mezi lékaři a výzkumníky jsou využívány „The Research Diagnostic Criteria for the Temporomandibular Disorders (RDC/ TMD)“. Na základě výše zmíněných faktů bychom rádi vytvořili EHR zaměřenou na TMD, která by vedla ke vzniku systému pro podporu rozhodování o terapii u TMD.

2. Metody

2.1 MUDR EHR, Stomatologická znalostní databáze a MUDRLite

EuroMISE Centrum má dlouhou tradici ve výzkumu a vývoji EHR-S systémů, která začala v roce 2000. Tento výzkum vedl k vyvinutí pilotní EHR aplikace nazvané MUDR (Multimedia Distributed EHR). Hlavní požadavek na tento systém bylo ukládání strukturovaných dat v kombinaci s volným textem s možností dynamického rozšiřování a modifikací nastavení sbíraných atributů beze změny struktury databáze. MUDR EHR byla založena na třívrstvé architektuře, která se skládá z vrstvy databáze, aplikační vrstvy a uživatelského rozhraní. Soubor definic sbíraných atributů a relací mezi nimi je ukládán v přímé grafické struktuře nazvané „knowledge base“.

Pro účely našeho projektu byly využity modelovací nástroje, které byly vyvinuty jako součást MUDR EHR k vytvoření Stomatologické znalostní databáze. Tato znalostní báze byla vyvinuta ve spolupráci IT specialisty se stomatologi v letech 2004 a 2005 a zahrnuje asi 1000 konceptů, které pokrývají všechny možné situace, se kterými se můžeme setkat v zubním statusu pacienta a také i základní data o tomto pacientovi.

Jelikož byl MUDR EHR navržen pro potřebu větších prostředí, což byl v některých případech limitující faktor, byl vyvinut jiný EHR-S systém v EuroMISE Centru v roce 2005. Tento EHR-S byl související s MUDR EHR a proto byl nazván MUDRLite [12]. Byl navržen s důrazem na potřebu menších praxí. Architektura MUDRLite je založena na dvouvrstvé struktuře. První je relační databáze (MS SQL server) a druhá je vrstva uživatelského rozhraní MUDRLite (MUDRLite UI). Schéma databáze koresponduje s jednotlivými potřebami cílové oblasti, a proto se výrazně mění v různých prostředích. Jádro MUDRLite – MUDRLite Interpreter – je schopno zacházet s rozličnými schématy databáze. Tato vlastnost často zjednodušuje možnost importu starších dat uložených v jiných databázích nebo souborech. Vizuální aspekty stejně jako chování MUDRLite UI je kompletně popsáno pomocí XML. Konečný uživatel může vidět soubor formulářů s různými ovladači pomocí odpovídajících XML elementů.

MUDRLite funguje jako překladatel povelů, zpracovává instrukce zašifrované v MLL (MUDRLite Language) [13] založeném na XML a pracuje s vrstvou databáze, stejně jako s vizuálními aspekty a chováním uživatelského rozhraní MUDRLite. Pilotní aplikace byla připravena pro stomatologii [8] a v dalším kroku pro TMD, což bude popsáno později.

2.2 DentCross komponenta, ASR

EHR systémy MUDR a MUDRLite byly vhodné pro všeobecné lékaře. Ve stomatologii je obvyklé mít informaci uloženou a prezentovanou v grafické formě. Z tohoto důvodu a pro lepší přijetí uživatelem, byla vyvinuta vylepšená komponenta reprezentující zubní kříž, která je jasně organizovaná v grafické podobě. DentCross komponenta je implementována jako samostatně stojící knihovna DentCross.dll, zcela vyvinuta pro .NET Framework platform. Uživatelem definovaná komponenta je vložena pomocí uživatelského elementu MLL (pojmenovaný „custom“) s následujícími povinnými atributy: „dll“ specifikuje jméno celku, kde je komponenta implementována a „class“ specifikuje jméno hlavní třídy zahrnuté komponenty.

Uživatelské rozhraní pro sběr TMD dat je uživatelský formulář definovaný kompletně v MLL. Je integrován ve stejné aplikaci jako DentCross a takto vytváří úplný EHR systém pro stomatologii [14].

Skutečné interaktivity je dosaženo pomocí zapojení systému pro automatické rozpoznávání řeči a modulu „text to speech“. Systém pro automatické rozpoznávání řeči (ASR) je více popsán v [15] a [16].

2.3 RDC/TMD

Pro speciální část EHR zaměřenou na TMD jsme použili základní schéma pro odebrání anamnestických údajů. Pro klasifikaci TMD byly využity části RCD/TMD, které aplikují systém dvou os pro diagnostiku a klasifikaci pacientů s TMD [17]. První osa je rozdělena do tří skupin běžně se vyskytujících TMD:

1. svalové poruchy, zahrnující myofasciální bolest,
2. dislokace disku, s nebo bez redukce,
3. artralgie, artritidy a artrózy.

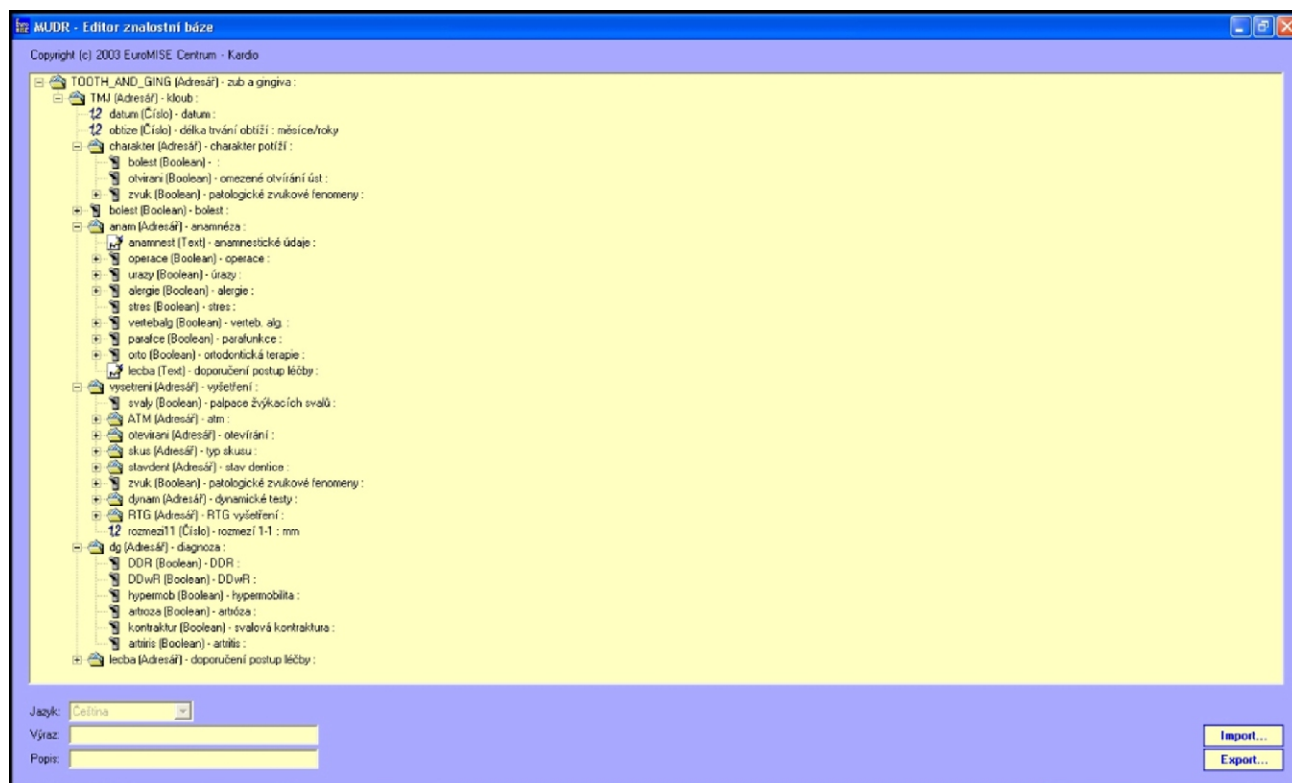
Druhá osa v této nomenklatuře používá soubor 31 otázek, které slouží k posouzení faktorů chování a psychologických a psychosociálních faktorů. My jsme tyto vyšetřované faktory modifikovali a vytvořili jsme náš vlastní dotazník pro posouzení kvality života.

3. Výsledky

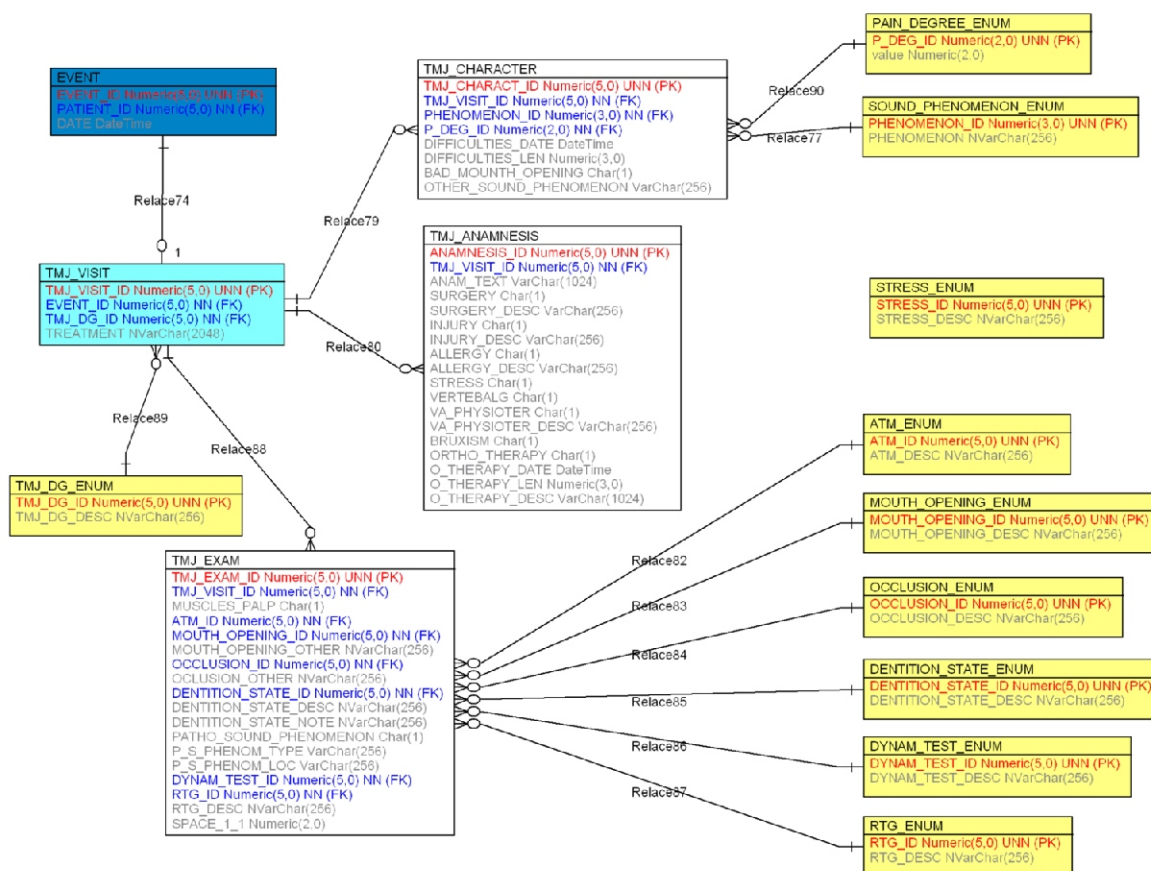
Nástroj znalostního modelování pro MUDR EHR aplikaci byl využit pro vytvoření speciální TMD znalostní databáze. Tato znalostní báze obsahuje více než 80 konceptů (Obr.1) a rozšiřuje dřívější Stomatologickou znalostní bázi.

Za účelem vytvoření MUDRLite formulářů musely být transformovány koncepty z TMD znalostní báze do modelu relační databáze (Obr. 2). Uživatelské rozhraní a aplikace pro TMD byly zakódovány do MLL v MUDRLite EHR a tím byly vytvořeny uživatelské formuláře propojené s vrstvou databáze.

Základní část našeho interaktivního EHR ve stomatologii je složena z DentCross komponenty s ASR systémem, která je postavena na základech vycházejících ze stomatologické znalostní báze a využívá MUDRLite EHR jako hostitelský systém.



Obr. 1. TMD znalostní báze.



Obr. 2. Datový model pro TMD EHR a jeho strom.

Specializovaná TMD EHR obsahuje koncepty definované v TMD znalostní bázi a je integrována do uživatelského rozhraní společně s DentCross komponentou. Aplikace se spouští z hlavního okna Integrated Dental-TMJ EHR (DentCross komponenta, ASR, TTS a TMD část) a má své vlastní uživatelské rozhraní.

Tato aplikace obsahuje 3 hlavní části (Obr. 3). V první části můžeme najít základní data pacienta (jméno, rodné číslo, datum narození atd.), což pomáhá v orientaci. Druhá část je seznam jednotlivých událostí. Ty jsou seřazeny podle datumů a otevřením těchto datumů se můžeme dostat k záznamům, které jsou spojeny přesně s touto návštěvou pacienta. Třetí záznamová část se skládá ze subjektivních problémů pacienta, anamnestických dat, výsledků různých vyšetření, stanovení diagnózy a doporučení v terapii. Tato složka je nejvíce důležitá a představuje všechna data reprezentující jednotlivého pacienta.

Je také velice důležité znát fakt, jaké obtíže vedly pacienta k návštěvě specialisty a jaké jsou celkové osobní pocity pacienta. Lékař může z tohoto pacientova popisu rozpoznat mnoho velice důležitých informací. Můžeme zaznamenat typ a charakter obtíží, jejich trvání a bližší popis. V anamnestické sekci můžeme zaznamene-

nat základní osobní anamnestická data a různé faktory podstatné pro TMD (jako např. parafunkce, stres, ortodontická terapie). Pravděpodobně nejdůležitější je sekce s různými typy vyšetření. Vyšetření u TMD se rozděluje na pohled (deformity, stav chrupu, otvírání úst (Obr. 4 atd.), pohmat (žvýkací svaly, temporomandibulární kloub sám o sobě) a poslech (zvukové fenomény). Poslední složky, které hrají 4. zásadní roli ve vyšetřování TMD, jsou rentgenové snímky (ortopantomografický snímek, počítačová tomografie (Obr. 5)) a magnetická resonance. Popisy a zhoto-vení těchto snímků mohou být také zaznamenány. Ke klasifikaci diagnóz bylo využito výše zmíněné RDC/TMD. Je možné také zaznamenat doporučení v terapii onemocnění, které se může měnit na konci každé návštěvy. Tato aplikace tedy slouží jako komplexní zdroj informací pro léčbu TMD a je součástí celého systému pro podporu rozhodování v terapii TMD, která nakonec vede k finálnímu terapeutickému plánu:

1. záznam do interaktivní DentCross komponenty – stomatologické vyšetření a anamnestická data,
2. nácvik otvírání úst v ose, bez deviací,
3. stabilizační dlahy
4. analýza pomocí počítačového obličejového oblouku (Obr. 6),
5. protetická terapie,

6. dokončení ošetření a měkká dlahy,
7. opětovná kontrola s odstupem a hodnocení ošetření pacientem

4. Diskuze

Prezentovaná DentCross komponenta, obzvláště se specializovanou částí pro TMD a ASR systémem, se jeví jako velice vhodná pro oblast stomatologie. Její grafická podoba dovoluje výborný přehled i během ošetření pacienta. Použití hlasových povelů a systému ASR umožňuje kontrolu ukládaných dat bez přítomnosti další osoby. Další výhodou je možnost záznamu dat bez tradiční kontroly počítače pomocí myši a klávesnice. To je velice podstatné především ve stomatologii, ale i v celé medicíně obecně. Ruka s rukavicí ošetřujícího je v častém kontaktu s potenciálně infekčním materiálem. DentCross komponenta je používána ve stomatologických ordinacích na Dětské stomatologické klinice, Fakultní nemocnici v Motole a byla také použita na Oddělení biologické analýzy, Kriministického ústavu v Praze pro identifikaci lidských ostatků [8].

Obr. 3. TMD EHR aplikace.

Specializovaná TMD část Integrated Dental – TMJ EHR rozšiřuje stávající EHR systémy ve stomatologii. Skládá se ze základních anamnestických dat, dat z proběhlých vyšetření, stanovení diagnózy a doporučení v další terapii. Tato aplikace byla použita ve specializované poradně ve Fakultní nemocnici v Motole pro terapii TMD a byla součástí systému pro podporu rozhodování v terapii TMD.

V budoucnu se naše práce soustředí na zlepšení Integrated Dental-TMJ EHR a tvorbu automatického systému, který na základě anamnestických znaků a symptomů stanoví diagnózu a připraví terapeutický plán. To by mělo zjednodušit a zlepšit celý terapeutický proces multifaktoriálně podmíněných TMD.

Poděkování

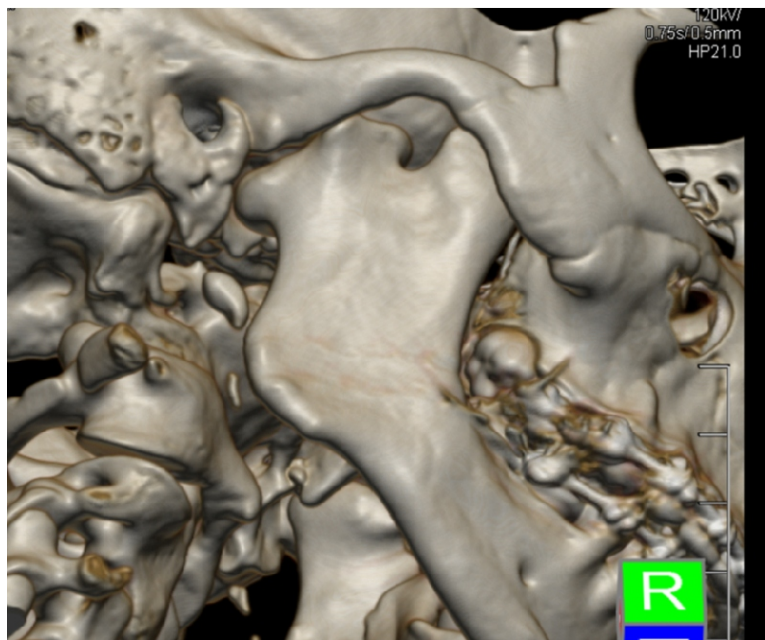
Tato práce byla podpořena z Grantu 9991-3 IGA MZČR, projektem FI-IM5/142, projektem 1M06014 MŠMT ČR a SVV-2010-265513.

Literatura

- [1] Shabbir S.A., Ahmed L.A., Sudhir R.R., Scholl J., Li Y.C., Liou D.M.: Comparison of documentation time between an electronic and a paper-based record system by optometrist at an eye hospital in south India: A time-motion study. *Comput Methods Programs Biomed*, 2010 May 10.
- [2] Liu D., Wang X., Pan F., Yang P., Xu Y., Tang X., Hu J., Rao K.: Harmonization of health data at national level: a pilot study in China. *Int J Med Inform*, 2010 Jun; 79(6): 450-8.
- [3] Los R.K., van Ginneken A.M., van der Lei J.: OpenSDE: a strategy for expressive and flexible structured data entry. *Int J Med Inform* 2005; 74:481-490.
- [4] van Ginneken A.M.: The computerized patient record: balancing effort and benefit. *Int J Med Inf* 2002; 65:97-119.
- [5] Bakker A.: Access to EHR and access control at a moment in the past: A discussion of the need and an exploration of the consequences. *Int J Med Inform* 2004; 73:267-270.
- [6] Blobel B.: Advanced EHR Architecture Promises or Reality. *Methods Inf Med* 2006; 1: 95-101.
- [7] Grasso M.A.: Automated speech recognition in medical applications. *MD Computing* 1995; 12: 16-23.
- [8] Zvarova J., Dostalova T., Hanzlicek P., Teuberova Z., Nagy M., Pies M., Seydlova M., Eliasova H., Simkova H.: Electronic health record for forensic dentistry. *Methods Inf Med* 2008; 47: 8-13.



Obr. 4. Temporomandibulární kloub – sběr dat. Měření rozsahu otvírání úst.



Obr. 5. Temporomandibulární kloub – 3D-CT zobrazení počítačové analýzy. Na tomto obrázku je prezentována ankylóza temporomandibulárního kloubu.

Bite Fork: Position 2			
Right Joint		Left Joint	
HCN/HCI	05.7 (CE)	HCN/HCI	-0.9 (CE)
Bennett	04.0	Bennett	04.0
Iss	00.0	Iss	00.0
Shift Angle	-17.9	Shift Angle	-20.0
Anterior Guidance			
Right	18.5		
Middle	14.7		
Left	16.0		
		Exit	Reset

Obr. 6. Temporomandibulární kloub – analýza počítačových zubním obloukem.

- [9] Di Fabio R.P.: Physical therapy for patient with TMD: a descriptive study of treatment, disability and health status. *J Orofac Pain*. 1998;12:124-135.
- [10] Pedroni C.R., De Oliveira A.S., Guaratini M.I.: Prevalence study of signs and symptoms of temporomandibular disorders in university students. *J Oral Rehabil*. 2003; 30: 944-950.
- [11] Taylor M., Suvinen T., Reade P.: The effect of Grade IV distraction mobilization on patients with temporomandibular pain-dysfunction disorder. *Physiotherapy Theory and Practice*. 1994;10:129-136.
- [12] Hippmann R., Dostalova T., Zvarova J., Nagy M., Seydlova M., Hanzlicek P., Kriz P., Smidl L., Trmal J.: Voice supported electronic health record for temporomandibular joint disorders. *Methods of information in medicine*, 2010; 49:168-172
- [13] Spidlen J., Hanzlicek P., Zvarova J.: MUDRLite – health record tailored to your particular needs. *Studies in health technology and informatics* 2004; 105: 202-9.
- [14] Hippmann R., Dostalova T., Zvarova J., Nagy M., Seydlova M., Hanzlicek P., Kriz P., Smidl L., Trmal J.: Voice supported electronic health record for temporomandibular joint disorders. *Methods of information in medicine*, 2010; 49:168-172.
- [15] Nagy M., Hanzlicek P., Zvarova J., Dostalova T., Seydlova M., Hippmann R., Smidl L., Trmal J., Psutka J.: Voice-controlled Data Entry in Dental Electronic Health Record. *Studies in Health Technology and Informatics* 2008; 136: 529-534.
- [16] Psutka J., Muller L., Matousek J., Radova V.: *We speak Czech with a computer* Praha, Academia, 2006.
- [17] Ali H.N.: Diagnostic criteria for temporomandibular joint disorders: a physiotherapist's perspective. *Physiotherapy*. 2002; 88:421-426.

Kontakt**MUDr. Radek Hippmann**Dětská stomatologická klinika 2. LF UK
a FN Motol

V úvalu 84

150 06 Praha 5

e-mail: r.hip@seznam.cz

Autoimunitní thyroditida – porovnání vyšetřujícího lékaře a automatické klasifikace, vyhodnocení variability lékařů

Š. Holinka¹, D. Smutek¹

¹3. interní klinika, 1. lékařská fakulta Univerzity Karlovy v Praze

Školitel: Doc. MUDr. Ing. Daniel Smutek, Ph.D.

Souhrn

Porovnání úspěšnosti texturního klasifikátoru a vyšetřujícího lékaře (radiologa) při diagnóze autoimunitní thyroditidy ze sonografického obrazu snímaného v B-módu. Určení inter- a intrapersonální variability lékařů.

Datový soubor obsahující 161 vyšetřovaných subjektů byl rozdělen do tří skupin dle celkového klinického vyšetření: normální – H (healthy); hraniční stav – BS (border state); autoimunitní thyroditida AT. Následně byl soubor čtyřmi vyšetřujícími lékaři a Bayesovským klasifikátorem, založeným na texturních příznacích, hodnocen do těchto skupin.

Dva lékaři dosáhli vyšší úspěšnosti při hodnocení subjektů z normální skupiny (74,4 % a 83,3 %) a dva lékaři hodnotili lépe subjekty s autoimunitní thyroditidou (59,0 % a 77,4 %). Klasifikátor dosáhl relativně vysoké a vyrovnané úspěšnosti pro obě tyto skupiny (100,0 % pro normální a 87,5 % pro thyroditidu). Rozdílný úspěch jednotlivých lékařů při hodnocení subjektů vyústil ve vyšší interpersonální variabilitu, tedy nízkou shodu mezi nimi. V intrapersonální variabilitě jednotlivých lékařů nebyl nalezen významný rozdíl.

Vzhledem ke slabé shodě mezi vyšetřujícími lékaři při diagnostice autoimunitní thyroditidy ze sonografických obrazů a vysoké a vyrovnané úspěšnosti klasifikátoru se zdá jako nejvýhodnější pro stanovení konečné diagnózy kombinace automatické klasifikace obrazů a klinických zkušeností lékařů.

Klíčová slova: štítná žláza; autoimunitní thyroditida; sonografický obraz; B-mode sonografie; texturní analýza; počítačem podporovaná diagnóza; interpersonální variabilita; intrapersonální variabilita; koeficient Kappa; vážený koeficient Kappa

1. Úvod

Autoimunitní thyroditida je jednou z nejčastějších onemocnění štítné žlázy. Jedná se o chronický zánět parenchymu [1]. Zánět žlázy způsobuje v její struktuře difúzní změny, které lze odhalit pomocí ultrazvukového vyšetření, jakožto nejrozsáhlejší diagnostické zobrazovací metody pro toto onemocnění. Ultrazvuk se tedy dnes používá jako doplňková diagnostická metoda vedle laboratorního vyšetření krve a cytologického vyšetření po punkci štítné žlázy. Nicméně hodnocení difúzních změn parenchymu ze sonografických obrazů v praxi je složité [2], [3]. Provádí se kvalitativní hodnocení na základě velikosti žlázy, perfúze, struktury a echogenity parenchymu závislé na klinických zkušenostech vyšetřujícího lékaře, tedy bez jakýchkoliv kvantitativních ukazatelů.

Takovéto subjektivní hodnocení textury sonografických obrazů je hlavním důvodem, proč se vyvíjejí počítačem podporované diagnostické metody. Jednou z nich je semi-automatická diagnostika autoimunitní thyroditidy ze sonografických obrazů, pro níž v minulých pracích byly nalezeny vhodné texturní příznaky [4], [5] a byl navržen úspěšný klasifikátor [6]. Počítačem podporovaná diagnostika má obecný význam i v dalších oblastech při detekci různých onemocnění, například screeningová mamografie [7] nebo pletysmografie [8].

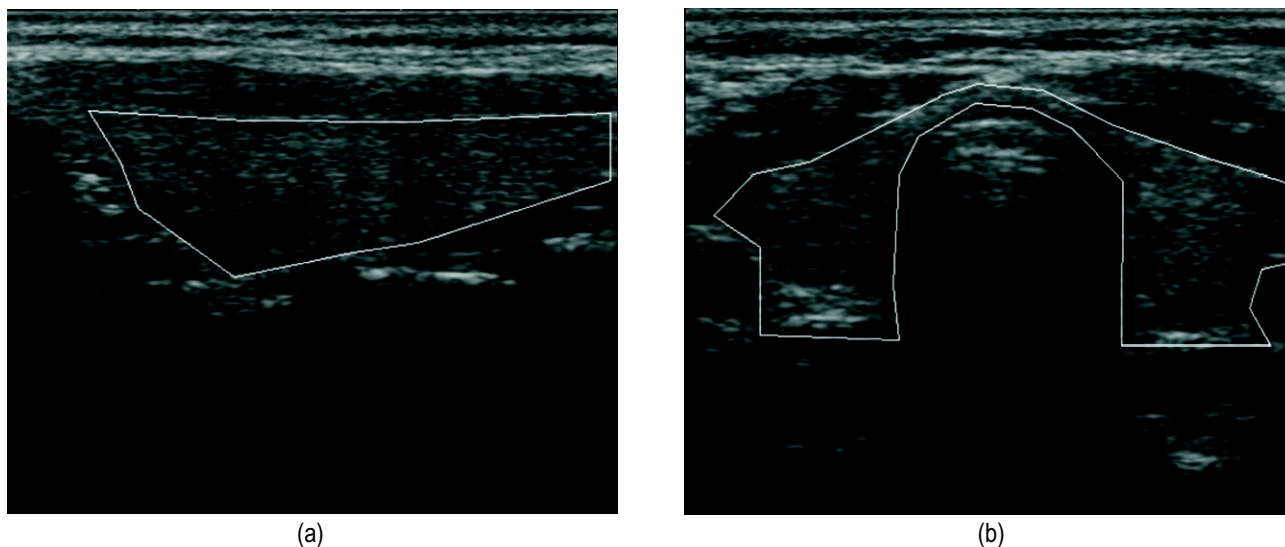
V tomto příspěvku se zabýváme porovnáním výsledku vyšetřujících lékařů (endokrinologů) při diagnostice autoimunitní thyroditidy ze sonografických obrazů v B-módu s výsledky získanými klasifikátorem popsáním v [6]. Dále byla získaná data vyhodnocena v rámci inter- a intrapersonální variability lékařů.

2. Data a metody

Použitými daty byla sada sonografických obrazů snímaných v B-módu získaných od 161 subjektů vyšetřených na 3. interní klinice (endokrinologie), u nichž bylo indikováno ultrazvukové vyšetření. Subjekty byly průběžně zařazovány do studie po tři měsíce, přičemž případy s lokálními změnami, jako jsou například uzly, byly vyřazeny. Pro každý subjekt bylo nasnímáno 4 až 8 sonografických obrazů v podélném a příčném řezu (Obr. 1) pomocí ultrazvukového systému EnVisor M2540A, vybaveného 8MHz lineární sondou. Obrazy byly uloženy ve formátu DICOM.

Data byla rozdělena do tří skupin podle diagnózy potvrzené celkovým klinickým vyšetřením: H (healthy) – normální žláza, 26 subjektů; BS – (borderline state) hraniční stav mezi normální a difúzně změněnou tkání, 14 subjektů; AT – autoimunitní thyroditida, 121 subjektů.

Čtyři lékaři (A, B, C, D), endokrinologové s minimálně desetiletou praxí ve vyšetřování štítné žlázy ultrazvukem, vyhodnocovali anonymizované sonografické obrazy pomocí webové aplikace, vyvinuté přímo pro tento účel, bez znalosti předchozí diagnózy nebo ostatních klinických výsledků. Lékaři měli za úkol každý zobrazený subjekt v aplikaci zařadit na základě obrazů do jedné ze skupiny (H, BS, AT a „Nehodnoceno“). Čtvrtá skupina „Nehodnoceno“ byla přidána pro případy, kdy při hodnocení vyšetřující lékař nebyl schopen zařadit subjekt do jedné ze skupin na základě dostupných obrazů. Časový limit na hodnocení nebyl stanoven. Lékaři celý úkol provedli během tří týdnů, přičemž u jednoho subjektu strávili v průměru 5 sekund. Celkově byl každý subjekt hodnocen každým lékařem třikrát (ve třech hodnotících kolech).



Obr. 1. Snímané sonografické obrazy štítné žlázy s vyznačenými oblastmi zájmu: (a) podélný řez tkáně s difúzními změnami; (b) příčný řez normální tkáně.

Současně byla data vyhodnocena klasifikátorem (CL) zkonstruovaným a testovaným v minulém projektu [6]. Klasifikátor je založen na Bayesovské rozhodovací teorii a využívá texturní příznaky navržené Muzzolinima Haralickem [9], [10]. Protože konkrétní ultrazukový přístroj poskytuje konkrétní numerické hodnoty texturních příznaků, jak bylo poznamenáno v [5], nebylo možné využít původní trénovací data. Proto byl dostupný soubor dat rozdělen na trénovací a nezávislou testovací množinu (viz Tabulka 1). Dále byl proveden nový výběr příznaků, přičemž jako nejvhodnější příznak z testované skupiny byl vybrán Muzzoliniho texturní příznak reprezentující obrazový gradient [9].

Nejdříve byla porovnávána samotná úspěšnost lékařů a klasifikátoru při hodnocení subjektů do správných skupin. Poté byla zkoumána inter- a intrapersonální variabilita lékařů pomocí koeficientu Kappa [11], [12]. Hodnoty Kappa obvykle nabývají hodnot z intervalu $<0,1>$ a jsou interpretovány jako slabá shoda pro $\kappa = 0$ a jako úplná shoda pro $\kappa = 1$. Přesněji, interpersonální variabilita byla vyhodnocena na skupině čtyř lékařů

zvlášť pro každé hodnotící kolo pomocí koeficientu vícenásobná Kappa κ_m a vážená kapa κ_w s kvadratickými váhami a 95% intervalem spolehlivosti. Vážená kapa κ_w umožňuje vyhodnotit závažnost neshody lékařů při hodnocení do více než dvou konečných diagnostických skupin (H, BS, AT) [12], přičemž velikost difference skupin je ohodnocena vahami. Intrapersonální variabilita byla kvantifikována pomocí koeficientu vícenásobná Kappa κ_m mezi třemi hodnotícími koly každého lékaře.

3. Výsledky

3.1 Porovnávací studie

Úspěšnosti při vyhodnocování dat lékaři (A, B, C, D) a klasifikátorem (CL) jsou shrnuty v Tabulce 1. Předpokladem bylo, že hodnocení každého subjektu je nezávislý náhodný jev. Proto celkové množství hodnocených subjektů pro každého lékaře je 483, což odpovídá 161 subjektům nezávisle hodnoceným ve třech kolech. Úspěšnost klasifikátoru byla vyhodnocena na nezávislé testovací množině (52 subjektů), protože zbylá část dat byla použita pro natrénování klasifikátoru.

Opakování hodnocení dat klasifikátorem nebylo třeba díky 100% reprodukcibilitě klasifikátoru.

Úspěch hodnocení je zobrazen jako absolutní počet správně ohodnocených subjektů i jako relativní údaj, vyjadřující poměr počtu správně ohodnocených subjektů a celkového počtu hodnocených subjektů v dané skupině. Například pro lékaře A platí:

$$\begin{aligned} H: & 33 / (3 \cdot 26) = 0,423 \\ BS: & 28 / (3 \cdot 14) = 0,667 \\ AT: & 214 / (3 \cdot 121) = 0,590 \end{aligned}$$

Sloupec „velká chyba“ vyjadřuje subjekt ze skupiny H vyhodnocený jako skupina AT a naopak. Sloupec „malá chyba“ vyjadřuje nesprávné hodnocení skupiny BS do H nebo AT a naopak. Relativní malá a velká chyba je poměrem počtu chybných hodnocení oproti celkovému počtu všech provedených hodnocení subjektů lékařem ($3 \cdot 161$) či klasifikátorem (52).

Tab. 1. Dostupná data.

	skupina H	skupina BS	skupina AT
trénovací množina	18	10	81
nezávislá testovací množina	8	4	40

Tab. 2. Úspěšnost při hodnocení subjektů lékaři (A,B,C,D) a klasifikátorem (CL) do tří daných skupin H, BS, AT. Úspěšnost je vyčíslena jako absolutní počet správných hodnocení i jako relativní hodnota v závorce.

Hodnotitel	Úspěch v daných skupinách			Celkový úspěch	Velká chyba	Malá chyba	Nehodnocen
	H	BS	AT				
A	33 (0,423)	28 (0,667)	214 (0,590)	275 (0,569)	16 (0,033)	190 (0,393)	2
B	58 (0,744)	22 (0,524)	173 (0,477)	253 (0,524)	47 (0,097)	165 (0,342)	18
C	13 (0,167)	25 (0,595)	281 (0,774)	319 (0,660)	13 (0,027)	151 (0,313)	0
D	65 (0,833)	16 (0,381)	139 (0,383)	220 (0,455)	75 (0,155)	169 (0,350)	19
CL	8 (1,000)	4 (1,000)	35 (0,875)	47 (0,904)	2 (0,038)	3 (0,058)	0

Tab. 3. Kontingenční tabulky zobrazují výsledky hodnocení lékaři a klasifikátorem (leave-one-out metoda použita na trénovací množinu a hodnocení nezávislé testovací množiny). Sloupce kontingenčních tabulek reprezentují správnou skupinu, řádky reprezentují rozhodnutí lékařů a klasifikátoru.

lékař A					lékař B				
	H	BS	AT	Nehodnocen		H	BS	AT	Nehodnocen
H	33	38	7	0	H	58	18	1	1
BS	9	28	5	0	BS	16	22	4	0
AT	9	138	214	2	AT	46	127	173	17

lékař C					lékař D				
	H	BS	AT	Nehodnocen		H	BS	AT	Nehodnocen
H	13	53	12	0	H	65	13	0	0
BS	1	25	16	0	BS	24	16	1	1
AT	1	81	281	0	AT	75	131	139	18

klasifikátor - LOO na trénovací množině				klasifikátor – nezávislá testovací množina			
	H	BS	AT		H	BS	AT
H	18	0	0	H	8	0	0
BS	3	6	1	BS	0	4	0
AT	5	8	68	AT	2	3	35

Výsledky hodnocení lékařů a klasifikátorem jsou znázorněny i pomocí kontingenčních tabulek, viz Tabulka 3.

3.2 Inter- a intrapersonální variabilita

Výsledky hodnocení inter- a intrapersonální variability lékařů shrnují Tabulky 4, 5 a 6. Sloupec „N“ vyskytující se v následujících tabulkách označuje počet hodnocených subjektů mimo ty, které byly označeny lékaři jako nehodnocené.

Rozlišujeme tři hodnotící kola (k.1-3), jednotliví lékaři ohodnotili všech 161 subjektů v každém. Interpersonální variabilita (Tabulka 2) byla vyčíslena jednotlivě pro každé ze tří kol, přičemž vícenásobná Kappa κ_m reprezentuje celkovou shodu mezi lékaři. Zároveň je v tabulce vyčíslen počet správně ohodnocených subjektů, u kterých se shodli všichni lékaři. Vážená Kappa ukazuje detailní závislosti u vytvořených párů lékařů.

Hodnoty ukazuje Tabulka 5. Znázorněna je shoda lékařů při hodnocení do jednotlivých skupin vypočítaná jako počet správně hodnocených subjektů do skupiny oběma lékaři v páru dělený celkovým počtem subjektů hodnocených do této skupiny oběma lékaři [10], [11].

Tab. 4. Interpersonální shoda lékařů (A,B,C,D). Vícenásobná Kappa κ_m pro jednotlivá hodnotící kola.

Hodnotitel	N	Shoda všech lékařů	κ_m
A,B,C,D (k.1)	152	63 (0.416)	0.448
A,B,C,D (k.2)	154	39 (0.253)	0.264
A,B,C,D (k.3)	147	31 (0.211)	0.258

Tab. 5. Interpersonální shoda mezi páry lékařů (A,B,C,D). Shoda ve skupinách H, BS a AT a vážená Kappa κ_m jednotlivě pro tři hodnotící kola.

Pár hodnotitelů	N	Shoda ve skupině			κ_m	95% interval spolehlivosti
		H	BS	AT		
A&B (k.1)	156	0,593	0,536	0,795	0,670	0,622 - 0,777
A&B (k.2)	158	0,333	0,465	0,707	0,544	0,453 - 0,635
A&B (k.3)	151	0,585	0,565	0,726	0,647	0,667 - 0,737
A&C (k.1)	160	0,533	0,682	0,823	0,698	0,614 - 0,781
A&C (k.2)	161	0,286	0,593	0,790	0,535	0,422 - 0,648
A&C (k.3)	160	0,174	0,474	0,754	0,499	0,400 - 0,598
A&D (k.1)	155	0,480	0,442	0,824	0,673	0,594 - 0,751
A&D (k.2)	155	0,281	0,397	0,720	0,544	0,462 - 0,627
A&D (k.3)	153	0,520	0,450	0,600	0,569	0,485 - 0,653
B&C (k.1)	156	0,429	0,536	0,823	0,675	0,570 - 0,751
B&C (k.2)	158	0,167	0,212	0,671	0,463	0,383 - 0,543
B&C (k.3)	151	0,087	0,226	0,627	0,402	0,320 - 0,484
B&D (k.1)	152	0,706	0,455	0,779	0,717	0,634 - 0,800
B&D (k.2)	154	0,714	0,527	0,720	0,698	0,614 - 0,782
B&D (k.3)	147	0,755	0,605	0,675	0,710	0,629 - 0,792
C&D (k.1)	156	0,286	0,385	0,759	0,521	0,423 - 0,619
C&D (k.2)	155	0,139	0,180	0,618	0,388	0,306 - 0,471
C&D (k.3)	153	0,069	0,175	0,463	0,298	0,225 - 0,370

Tab. 6. Intrapersonální shoda lékařů (A,B,C,D). Vícenásobná Kappa κ_m pro jednotlivé lékaře.

Hodnotitel	N	Shoda ve všech kolech	κ_m
A	160	94 (0.588)	0.534
B	151	89 (0.589)	0.578
C	161	108 (0.671)	0.545
D	151	80 (0.530)	0.525

Intrapersonální variance (Tabulka 6) udává shodu při hodnocení každého subjektu ve třech kolech pro každého lékaře samostatně. Rozdílné hodnoty ve sloupci „N“ jsou dány vypuštěním subjektů označených alespoň jednou jako nehodnocené.

4. Diskuze a závěr

Ve srovnávací studii byla vyhodnocena úspěšnost hodnocení subjektů na základě sonografických obrazů štítné žlázy. Čtyři hodnotící lékaři vytvořili dva páry s podobnou úspěšností, a to pár AC a pár BD. Lékaři A a C dosáhli vyšší úspěšnosti při hodnocení do skupiny AT (59,0 % a 77,4 %), přičemž lékař C označil několik normálních subjektů (H) jako pozitivní na autoimunitní thyroditidu, tzn. hodnotil data s falešnou pozitivitou – 12 z 13-ti velkých chyb je nesprávné hodnocení subjektů ze skupiny H do skupiny AT.

Lékař A měl velké chyby vyrovnané – 7 z 16 je nesprávné hodnocení subjektů ze skupiny H do skupiny AT, zbývajících 9 chyb naopak (viz Tabulka 3).

Proti tomu lékaři B a D hodnotili s větší úspěšností do skupiny H (74,4 % a 83,3 %) a využili více možností označit některé subjekty jako neohodnotitelné (18 a 19 subjektů). Dosáhli také vyššího počtu velkých chyb (47 a 75), kdy subjekty ze skupiny AT byly nesprávně hodnoceny do skupiny H, tzn. s falešnou negativitou (viz Tabulka 3).

Klasifikátor byl ověřen metodou leave-one-out použitou na trénovací množinu, přičemž dosáhl relativně vysoké celkové úspěšnosti 84,4 % (viz Tabulka 3) při rozlišení třech zvolených skupin (H, BS, AT).

Úspěšnost hodnocení vyčíslená na nezávislé testovací množině byla vyrovnaná (100,0 % pro H; 100,0 % pro BS a 87,5 % pro AT). Obě dvě velké chyby reprezentují nesprávné hodnocení subjektů ze skupiny AT do skupiny H (viz Tabulka 3), což reprezentuje falešnou negativitu. Jak bylo zmíněno v úvodu, lékaři se při hodnocení subjektů rozhodují subjektivně podle struktury a echogenity parenchymu na sonografickém obrazu a podle nabytých klinických zkušeností, zatímco klasifikátor vyhodnocuje kvantitativní ukazatele z textury obrazu. Proto se zdá jako nejvýhodnější pro stanovení konečné diagnózy kombinace automatické klasifikace obrazů a klinických zkušeností lékařů.

Ve studii zabývající se inter- a intra personální variabilitou se ukázalo, že v prvním hodnotícím kole byla shoda mezi lékaři vyšší ($\kappa_m = 0,448$) než ve dvou následujících ($\kappa_m = 0,264$ a $0,258$), viz Tabulka 4. Důvodem je nejspíše to, že lékaři v prvním kole hodnotili data s větším soustředěním a zájmem než v kolech následujících.

V Tabulce 5 je zobrazena shoda mezi páry lékařů pomocí koeficientu κ_w , přičemž souhlas v párech tvořených lékařem C s lékaři B a D je nižší než v párech tvořených lékaři A, B, D. Důvodem je značná falešná pozitivita lékaře C (viz Tabulka 2) oproti vyšší falešné negativitě v hodnoceních prováděných lékaři B a D. Shoda mezi lékaři A, B a D se pohybuje v intervalu od střední po značnou. Nejlepší shoda se vyskytla v párech AB a BD, což je pravděpodobně způsobeno lékařem B, který hodnotil data s úspěšností v rozmezí mezi lékaři A a D (Tabulka 2).

Intrapersonální variabilitu vyčíslenou pomocí koeficientu vícenásobná Kappa (κ_m) ukazuje Tabulka 6, přičemž nebyl nalezen významný rozdíl mezi jednotlivými lékaři (κ_m : 0,534; 0,578; 0,545 a 0,525). Pokud uvažujeme hodnoty Kappa z intervalu $<0,1>$, kde $\kappa_m = 0$ znamená žádnou shodu a $\kappa_m = 1$ znamená dokonalou shodu, všichni lékaři ve svých hodnoceních dosáhli střední intra personální shody.

Naše příští studie bude zaměřena na hypotézu, že lékaři klinické zkušenosti podpořené rozhodnutím navrženého klasifikátoru by měli ještě zvýšit celkový úspěch při diagnostice autoimunitní thyroditidy ze sonografických obrazů štítné žlázy.

Poděkování

Autoři děkují endokrinologům Prof. MUDr. Michalu Krškovi CSc., MUDr. Janu Jiskrovi PhD a MUDr. Petru Suchardovi z 3. interní kliniky endokrinologie a metabolismu Univerzity Karlovy a VFN v Praze za jejich úsilí při hodnocení sonografických obrazů.

Tato práce je podporována Grantovou agenturou Univerzity Karlovy v Praze (GAUK 119607) a grantem SVV-2010-265 513.

Literatura

- [1] Warfatsky L., Ingbar S.H. : Disease of the thyroid. In Harrison's Principles of Internal Medicine, page 1712. McGraw-Hill, New York, 12th edition, 1991.
- [2] Simeone F.J., Daniel G.H., Müller P.R. et al. : High-resolution real-time sonography. Radiology, 155:431439, 1985.
- [3] Solbiati L., Volterrani L., Rizzatto G., et al. : The thyroid gland with low-uptake lesions: Evaluation by ultrasound. Radiology, 155:187196, 1985.
- [4] Šára R., Smutek D., Sucharda P., Svačina Š. : Systematic construction of texture features for Hashimoto's lymphocytic thyroiditis recognition from sonographic images. In S. Quaglini, P. Barahona, and S. Andreassen, editors, Artificial Intelligence in Medicine, LNCS, Berlin-Heidelberg, Germany. Springer, 2001.

- [5] Smutek D., Šára R., Sucharda P., Tjahjadi T., Švec M. : Image texture analysis of sonograms in chronic inflammations of thyroid gland. Ultrasound in Medicine and Biology, 29(11):1531 1543, 2003.
- [6] Holinka Š., Šára R., Smutek D. : Relation Between Structural Changes in B-mode Ultrasound Images of Thyroid Parenchyma and the Presence of Thyroid Antibodies in Blood Sample. Machine Graphics Vision, 18(1):67-82, 2009.
- [7] Georgian-Smith D., Moore R.H., Halpern E. : Blinded Comparison of Computer-Aided Detection with Human Second Reading in Screening Mammography. American Journal of Roentgenology, 189:11351141, 2007.
- [8] Paez S., Seiden D., Kiel M., Chediak A. : A comparison of human and computer apnea/hypopnea detection from respiratory waveforms derived by respiratory inductive plethysmography. Sleep Research, 25: 521, 1996.
- [9] Muzzolini R., Yang Y.-H., Pierson R. : Texture characterization using robust statistic. Pattern Recognition, 27(1):119134, 1994.
- [10] Haralick R.M., Sapiro L.G. : Computer and vision. vol. 1. Reading: Addison-Wesley Publishing Company, 1993: 453 506.
- [11] Fleiss J.L., Levin B., Paik M.C. : Statistical methods for rates and proportions. John Wiley, New York, third edition, 2003.
- [12] Kundel H.L., Polansky M. : Measurement of observer agreement. Radiology, 228(2):303308, August, 2003.

Kontakt

MUDr. Štěpán Holinka

3. Interní klinika 1. LF UK a VFN
U Nemocnice 1
128 08 Praha 2
e-mail: stepanholinka@gmail.com

Vztah zobecněné entropie a Bayesovské pravděpodobnosti chyby

M. Horáček¹

¹Centrum biomedicínské informatiky, Ústav informatiky AV ČR, v.v.i.

Školitelka: Prof. RNDr. Jana Zvárová, DrSc.

Souhrn

Cílem práce je zkoumat vztah mezi zobecněnou entropií diskrétní náhodné veličiny (tzv. f -entropií, třídou funkcí, zahrnující řadu indexů používaných při měření bio-diverzity) a minimální Bayesovskou pravděpodobností chyby při odhadu hodnoty této náhodné veličiny. Zejména studovat těsnost jejich vztahu. Morales a Vajda [1] nedávno zavedli míru zvanou průměrná nepřesnost (average inaccuracy), která kvantifikuje těsnost vztahu mezi aposteriorní Bayesovskou pravděpodobností chyby a mocninnými entropiemi. Tato míra je definována jako normalizovaný průměrný rozdíl horní a dolní meze aposteriorní Bayesovské pravděpodobnosti chyby za dané entropie. Tuto míru je možno zobecnit na jakoukoli striktně konkávní f -entropii a použít ji k vyjádření těsnosti vztahu této f -entropie k aposteriorní Bayesovské pravděpodobnosti chyby. Získaný vztah je však většinou poměrně složitý a není možné snadno analyticky porovnat průměrné nepřesnosti různých f -entropií.

Navrhujeme hladkou aproximaci dolní meze aposteriorní Bayesovské pravděpodobnosti chyby za dané f -entropie, jejíž použití zjednoduší formuli průměrné nepřesnosti. Ukázali jsme, že při použití této aproximace má kvadratická entropie nejtěsnější vztah k aposteriorní Bayesovské pravděpodobnosti chyby mezi f -entropiemi.

Kvadratická entropie má těsnější vztah k Bayesovské pravděpodobnosti chyby (ve smyslu popsaném v článku) než Shannonova entropie a další funkce příslušící do třídy f -entropií, jako např. Emlenův index, Ferreriho index, Goodův index a další.

Klíčová slova: zobecněná entropie, f -entropie, Bayesovská pravděpodobnost chyby, průměrná nepřesnost, mocninná entropie, kvadratická entropie Shannonova

va entropie, Emlenův index, Ferreriho index, Goodův index

1. Úvod

Nechť X a Y jsou diskrétní náhodné veličiny s hodnotami v konečném prostoru $X = (x_1, \dots, x_r)$, resp. $Y = (y_1, \dots, y_s)$. Nechť Z je náhodná veličina absolutně spojitá vzhledem k Lebesgueově míře s pravděpodobnostním rozdělením P_z na množině $Z \subseteq \mathbf{R}$. Budeme používat zjednodušené značení, např. $P(x) = P(X = x)$, pro sdružené rozdělení veličin X a Y $p(x, y) = P(X = x, Y = y)$, u podmíněného rozdělení X za dané hodnoty $Z = z$ bude $p(x|z)$ zastupovat $P(X = x|Z = z)$ atp., tj. argument vyjadřuje náhodnou veličinu, ke které se rozdělení vztahuje.

Nejistota spojená s náhodnou veličinou se tradičně měří pomocí Shannonovy entropie

$$H(X) = - \sum_{x \in X} p(x) \ln p(x)$$

s konvencí $0 \ln 0 = 0$. Tato nejistota je intuitivně spojena s minimální pravděpodobností chyby (apriorní Bayesovskou pravděpodobností chyby)

$$e(X) = 1 - \max_{x \in X} p(x)$$

při odhadu hodnoty X . Podmíněná Shannonova entropie

$$\begin{aligned} H(X|Z) &= \int_Z H(X|Z=z) dP_z(z) = \\ &= - \int_Z \sum_{x \in X} p(x|z) \ln p(x|z) dP_z(z) \end{aligned}$$

a minimální aposteriorní pravděpodobnost chyby (aposteriorní Bayesovská pravděpodobnost chyby)

$$e_B(X|Z) = 1 - \int_Z \max_{x \in X} p(x|z) dP_z(z)$$

k sobě mají podobný vztah. Fano [2] a Kovalevsky [3] byli mezi prvními, kdo studoval vztah mezi entropií a Bayesovskými pravděpodobnostmi chyby. Odvodili hodnoty dosažitelných mezi Shannonovy entropie $H(X)$ za dané apriorní Bayesovské pravděpodobnosti chyby $e(X)$. Protože jsou tyto meze ryze rostoucími funkcemi na intervalu $[0, (r-1)/r]$, jsou dobře definovány také meze apriorní Bayesovské pravděpodobnosti chyby za dané entropie. Poznatky Fana a Kovalevského byly později zobecněny na širší třídy entropií různými autory, jmenovitě Vajdou a Vaškem [4] na třídu všech Schur-konkávních entropií. Vajda a Vašek [4] také odvodili odpovídající meze pro podmíněné entropie a aposteriorní Bayesovskou pravděpodobnost chyby.

Ve své nedávné práci studovali Morales a Vajda [1] těsnost vztahu mezi apriorní, respektive aposteriorní Bayesovskou pravděpodobností chyby a třídou mocninných entropií. Odvodili vztahy pro průměrnou nepřesnost u těchto entropií a numerickými výpočty ukázali, že kvadratická entropie se zdá mít mezi mocninnými entropiemi výjimečné postavení tím, že má pro $r > 2$ nejtěsnější vztah k aposteriorní Bayesovské pravděpodobnosti chyby.

My se zabýváme podobným problémem u rodiny f -entropií, která zahrnuje značnou část z indexů používaných při měření biologické diverzity, které jsou využívány v biologických a medicínských oborech, například při výzkumu rakoviny (Maley a kol. [5]) či při studiu vztahů mezi různými populacemi (Chakraborty [6]). Tyto f -diverzity mimo jiné zahrnují také část mocninných entropií. Navrhujeme aproximaci dolní meze aposteriorní Bayesovské pravděpodobnosti chyby za dané f -entropie, která zjednoduší vzorec pro výpočet průměrné nepřesnosti. Poté analyticky ukážeme, že za této aproximace má kvadratická entropie mezi f -entropiemi nejtěsnější vztah k aposteriorní Bayesovské pravděpodobnosti chyby.

f -entropie

Zobecněná entropie, nazvaná f -entropie, byla navržena Zvárovou [7]. Zde stručně ukážeme, jak lze tuto funkci získat.

Informační divergenci

$$D(p(x) \| q(x)) = \sum_{x \in X} p(x) \ln(p(x) / q(x))$$

dvou diskrétních rozdělení $p(x)$ a $q(x)$ definovaných na jednom prostoru X zavedli Kullback a Liebler [8]. Entropií

$$H(X) = - \sum_{x \in X} p(x) \ln p(x)$$

a vzájemnou informaci

$$I(X; Y) = \sum_{x \in X} \sum_{y \in Y} p(x, y) \ln \frac{p(x, y)}{p(x)p(y)}$$

zavedl Shannon [9]. Tyto funkce splňují rovnosti

$$I(X; Y) = D(p(x, y) \| p(x)p(y))$$

$$H(X) = I(X; X) \quad (1)$$

a nerovnost

$$0 \leq I(X; Y) \leq H(X)$$

Obecnější třídu divergencí, nazvanou f -divergence, navrhnul Csiszár [10]. Nechť je $p(x)$ absolutně spojitě vůči $q(x)$, tj. $q(x) = 0 \Rightarrow p(x) = 0$, $x \in \chi$. Pak jsou f -divergence definovány vztahem

$$D_f(p(x) \| q(x)) = \sum_{x \in X} q(x) f\left(\frac{p(x)}{q(x)}\right)$$

pro jakoukoli reálnou funkci f , která je konvexní na intervalu $[0, \infty)$, ryze konvexní v $f(1)$, $f(1) = 0$ a $f(0) < \infty$ a je použita konvence $0f(0/0) = 0$. Poté můžeme stejně jako Zvárová [7] nahradit v definici vzájemné informace informační divergenci f -divergencí a zavést zobecněnou vzájemnou informaci vztahem

$$I_f(X; Y) = D_f(p(x, y) \| p(x)p(y)).$$

Nechť f je funkce splňující podmínky uvedené při definici f -divergence a nechť také $(f(x) - f(0))/x$ je funkce konkávní na intervalu $[1, \infty)$. f -entropie je pak definována vztahem

$$H_f(X) = I_f(X, X) =$$

$$\sum_{x \in X} p^2(x) f(1/p(x)) + f(0) \left(1 - \sum_{x \in X} p^2(x)\right).$$

Zvárová [7] ukázala, že f -entropie sdílejí mnoho vlastností s původní Shannonovou entropií, například za daného $r = \text{card}(\chi)$ je dosaženo nejmenší možné hodnoty (nulové entropie) f -entropie právě tehdy, když X je konstanta s pravděpodobností jedna, maximální hodnoty f -entropie je dosaženo, když X má rovnoměrné rozdělení na χ a také je splněna nerovnost

$$0 \leq I_f(X; Y) \leq H_f(X).$$

Další informace o vlastnostech a použití f -entropií je možno nalézt například v článku Zvárová, Vajda [11] a v práci Horáček [12].

Průměrná nepřesnost

Nechť $r = \text{card}(\chi) \geq 2$ a označme $\Delta_r = \{(p(x_1), \dots, p(x_r)) : p(x_i) \geq 0, \sum_{i=1}^r p(x_i) = 1\}$. Meze f -entropie za dané aposteriorní Bayesovské pravděpodobnosti chyby jsou zavedeny výrazy

$$H_f^+(e) = \max_{\{X: e(X)=e\}} H_f(X),$$

respektive

$$H_f^-(e) = \min_{\{X: e(X)=e\}} H_f(X)$$

tj. maximum (minimum) je vzato vzhledem k libovolné diskrétní distribuci $p(x)$ náhodné veličiny X na prostoru χ , která splňuje $e(X) = 1 - \max_{x \in \chi} p(x) = e$. Jelikož jakákoli f -entropie je Schur-konkávní funkcí na Δ_r (viz Horáček [12]), meze splňují

$$H_f^+(e) = H(1-e, e/(r-1), \dots, e/(r-1)) =$$

$$= (1-e)^2 f\left(\frac{1}{1-e}\right) + \frac{e^2}{r-1} f\left(\frac{r-1}{e}\right) + ef(0) \left(2 - e \frac{r}{r-1}\right)$$

když $e \in [0, (r-1)/r]$ a

$$H_f^-(e) = H(1-e, \dots, 1-e, 1-(1-e)k, 0, \dots, 0) =$$

$$= k(1-e)^2 f\left(\frac{1}{1-e}\right) + [1-k(1-e)^2] f\left(\frac{1}{1-k(1-e)}\right) +$$

$$+ f(0) [k(1-e)^2 + 1 - (1-k(1-e))^2]$$

když $e \in [(k-1)/k, k/(k+1)]$, $k \in \{1, \dots, r-1\}$.

Protože funkce $(f(x) - f(0))/x$ je konkávní na intervalu $[1, \infty)$ a tedy také funkce $x^2[f(x^{-1}) - f(0)]$ je konkávní na intervalu $[0, 1]$, není těžké ukázat, že horní mez $H_f^+(e)$ je konkávní na intervalu $[0, (r-1)/r]$ a dolní mez $H_f^-(e)$ je po částech konkávní na intervalech $[(k-1)/k, k/(k+1)]$, kde $k \in \{1, \dots, r-1\}$. Z konvexity f navíc plyne, že po částech lineární funkce

$$\Phi(t) = H_f\left(\frac{k-1}{k}\right) + \frac{t - \frac{k-1}{k}}{k+1 - \frac{k-1}{k}} \cdot$$

$$\left[h_f\left(\frac{k}{k+1}\right) - h_f\left(\frac{k-1}{k}\right) \right]$$

Kde $t \in [(k-1)/k, k/(k+1)]$, $k \in \{1, \dots, r-1\}$ je konvexní na intervalu $[0, (r-1)/r]$.

Horní mez f -entropie za dané *aposteriorní* Bayesovské pravděpodobnosti e_B chyby je definována jako

$$H_f^+(e_B) = \max_{\{(X,Z): e_B(X|Z)=e_B\}} H_f(X|Z),$$

tj. maximum je vzato vzhledem k libovolné diskrétní náhodné veličině X na prostoru \mathcal{X} a náhodné veličině Z , které splňují $e_B(X|Z) = e_B$. Dolní mez je definována analogicky. Protože meze f -entropie za dané *apriorní* Bayesovské pravděpodobnosti chyby jsou konvexní, respektive konkávní ve smyslu uvedeném výše, užití obdoby důkazu věty 4.1 z článku Morales a Vajda [1] vede k výsledku

$$\begin{aligned} H_f^+(e_B) &= (1-e_B)^2 f\left(\frac{1}{1-e_B}\right) + \\ &+ \frac{e_B^2}{r-1} f\left(\frac{r-1}{e_B}\right) + \\ &+ e_B f(0) \left(2 - e_B \frac{r}{r-1}\right) \end{aligned}$$

když $e \in [0, (r-1)/r]$ a dolní mez je konvexní, po částech lineární funkce shodnotami

$$H_f^-(e_B) = (1-e_B) f\left(\frac{1}{1-e_B}\right) + f(0) e_B$$

když

$$e_B = \frac{k}{k+1}, \quad k \in \{0, \dots, r-1\},$$

tj. $H_f^-(e_B) = \Phi(e_B)$.

Protože tyto meze jsou spojitými ryze rostoucími funkcemi na intervalu $[0, (r-1)/r]$, meze *aposteriorní* Bayesovské pravděpodobnosti chyby za dané f -entropie jsou také dobře definovány. Tyto meze označíme stejně jako Morales a Vajda [1]

$$\begin{aligned} e_{B,f}^+(H) &= \max_{H_f^-(e_B) \leq H} e_B \\ e_{B,f}^-(H) &= \max_{H_f^+(e_B) \leq H} e_B. \end{aligned}$$

Označme také

$$a_{r,f} = H_f\left(\frac{1}{r}, \dots, \frac{1}{r}\right)$$

a připomeňme, že mocninné entropie jsou definovány vztahem

$$H_\alpha(X) = \frac{1}{\alpha-1} \left(1 - \sum_{x \in X} p^\alpha(x)\right)$$

když $\alpha > 0$.

(Entropie H_1 je dodefinována limitně a odpovídá Shannonově entropii.) Mocninné entropie jsou f -entropie pro $\alpha = (0, 2]$, kvadratická entropie $H_2(X)$ je f -entropie s $f(x) = x-1$.

Morales a Vajda zavedli míru těsnosti vztahu mezi mocninnými entropiemi a *aposteriorní* Bayesovskou pravděpodobností chyby, kterou nazvali průměrná nepřesnost. Je definována vzorcem

$$AI_r(e_B | H_\alpha) = \frac{1}{a_{r,\alpha}} \cdot$$

$$\int_0^{a_{r,\alpha}} [e_{B,\alpha}^+(H) - e_{B,\alpha}^-(H)] dH.$$

Tato míra se dá přímočaře zobecnit na jakoukoli ryze konkávní f -entropii tak, že ve vzorci nahradíme $\alpha_{r,\alpha}$ za $\alpha_{r,f}$ a $e_{B,\alpha}$ za $e_{B,f}$.

Morales a Vajda také na základě numerických výpočtů vyzorovali, že kvadratická entropie se zdá mít mezi mocninnými entropiemi nejnížší průměrnou nepřesnost při odhadu *aposteriorní* Bayesovské pravděpodobnosti chyby. Kvůli složitému tvaru vzorce průměrné nepřesnosti je ale obtížné toto pozorování nějak analyticky ověřit. Navrhujeme proto postup, ve kterém využijeme aproximaci dolní meze $H_f^-(e_B)$,

při jejímž použití dostaneme přibližný, ale jednodušší tvar průměrné nepřesnosti. Aproximovaná dolní mez je tvaru

$$\tilde{H}_f^-(e_B) = (1-e_B) f\left(\frac{1}{1-e_B}\right) + f(0) e_B$$

na intervalu $[0, (r-1)/r]$ a je pro f -diverzity opět spojitá ryze rostoucí funkce a inverzní funkce $\tilde{e}_{B,f}^+(H)$ je tedy dobře definována. Tato dolní mez je také konvexní funkcí na intervalu $[0, (r-1)/r]$, v bodech $\{(k-1)/k, k: \{1, \dots, r-1\}\}$ nabývá stejných hodnot jako původní po částech lineární dolní mez, a nemůže tedy být větší než původní mez. Průměrnou nepřesnost pak aproximujeme vztahem

$$\tilde{AI}_r(e_B | H_f) = \frac{1}{a_{r,f}} \int_0^{a_{r,f}} [\tilde{e}_{B,f}^+(H) - \tilde{e}_{B,f}^-(H)] dH.$$

Ukázku mezi $H_f^+(e_B)$, $H_f^-(e_B)$ a $\tilde{H}_f^-(e_B)$ je možné vidět na Obrázku 1.

O takto přiblížené nepřesnosti je pak možné dokázat následující tvrzení.

Tvrzení 1 *Kvadratická entropie $H_2 = 1 - \sum_{x \in \mathcal{X}} p^2(x)$ má při použití aproximované průměrné nepřesnosti nejtěsnější vztah k *aposteriorní* Bayesovské pravděpodobnosti chyby mezi f -entropiemi, tj.*

$$\tilde{AI}_r(e_B | H_2) \leq \tilde{AI}_r(e_B | H_f)$$

pro jakoukoli f -entropii H_f a dané přirozené $r \geq 2$.

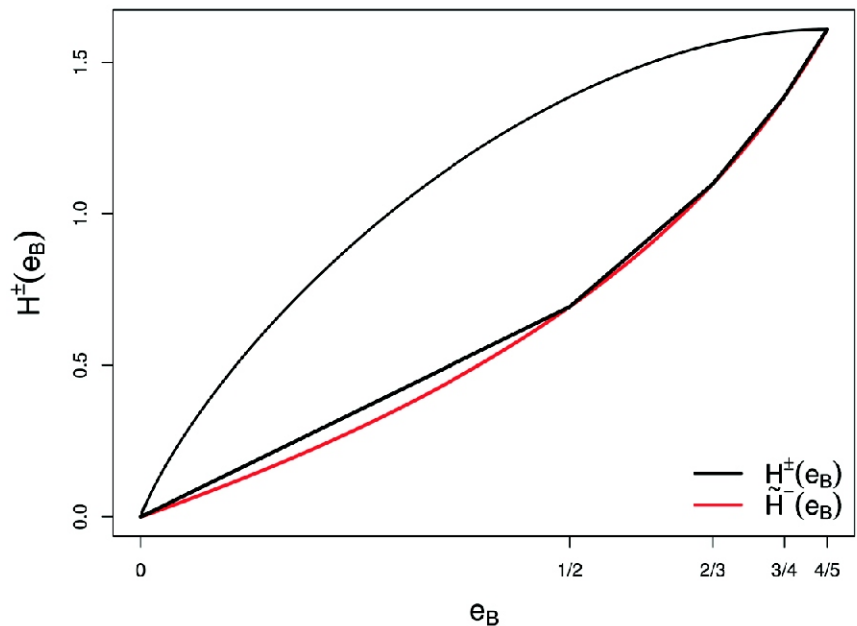
Důkaz. Připomeňme, že je-li H_f f -entropie, funkce f je konvexní na intervalu $[0, \infty)$ a ryze konvexní v bodě $f(1)$ s $f(1) = 0$. Uvažujme nejprve pouze funkce f splňující také $f(0)=0$. Z tohoto předpokladu plyne $f(x) > 0$ pro $x > 1$. Protože $H_f^+(0) = H_f^-(0)$ a také $H_f^+((r-1)/r) = H_f^-((r-1)/r)$, aproximovanou nepřesnost $\tilde{AI}_r(e_B, H_f)$ tj. normovaný obsah plochy mezi mezemi $\tilde{e}_{B,f}^+(H)$ a $\tilde{e}_{B,f}^-(H)$, můžeme vyjádřit také jako obsah plochy mezi mezemi $H_f^+(e_B)$ a $H_f^-(e_B)$ normovaný příslušnou konstantou.

$$\begin{aligned}
\tilde{A}I_r(e_B | H_f) &= \\
&= \frac{1}{a_{r,f}} \int_0^{a_{r,f}} [\tilde{e}_{B,f}^+(H) - e_{B,f}^-(H)] dH = \\
&= \frac{1}{a_{r,f}} \int_0^{\frac{r-1}{r}} [H_f^+(e_B) - \tilde{H}_f^-(e_B)] de_B = \\
&= \frac{r}{f(r)} \int_0^{\frac{r-1}{r}} (1-e_B)^2 f\left(\frac{1}{1-e_B}\right) + \\
&\quad + \frac{e_B^2}{r-1} f\left(\frac{r-1}{e_B}\right) - (1-e_B) \cdot \\
&\quad \cdot f\left(\frac{1}{1-e_B}\right) de_B = \frac{r}{f(r)} \int_0^{\frac{r-1}{r}} (1-e_B)^2 \cdot \\
&\quad \cdot f\left(\frac{1}{1-e_B}\right) + \frac{e_B^2}{r-1} f\left(\frac{r-1}{e_B}\right) - \\
&\quad - (1-e_B) f\left(\frac{1}{1-e_B}\right) de_B = \frac{r}{f(r)} \cdot \\
&\quad \cdot \int_0^{\frac{r-1}{r}} e_B \left[f\left(\frac{r-1}{e_B}\right) \frac{e_B}{r-1} - f\left(\frac{1}{1-e_B}\right) - \right. \\
&\quad \left. - (1-e_B) \right] de_B
\end{aligned}$$

Po transformaci $z=(r/r-1)e_B$ dostaneme

$$\begin{aligned}
\tilde{A}I_r(e_B | H_f) &= \frac{r}{f(r)} \int_0^1 \left(\frac{r-1}{r}\right)^2 \cdot \\
&\cdot \left[f\left(\frac{r}{z}\right) \frac{z}{r} - f\left(\frac{r}{r-zr+z}\right) \frac{r-zr+z}{r} \right] dz \\
&\quad (2).
\end{aligned}$$

Protože f je konvexní, $z \in (0,1]$ a $r \geq 1$, dostáváme nerovnosti



Obr. 1. Horní a dolní meze $H^\pm(e_B)$ a aproximovaná dolní mez $\tilde{H}^-(e_B)$ u Shannony entropie s $r = 5$.

$$\frac{f(r/z) - f(1)}{r/z - 1} \geq \frac{f(r) - f(1)}{r - 1}$$

$$\frac{f(r/z)}{r/z - 1} \geq \frac{f(r)}{r - 1}$$

$$\frac{f(r/z)z}{f(r)} \geq \frac{r-z}{r-1}$$

(3).

Podobně vztah

$$\frac{f[r/(z-zr+r)] - f(1)}{r/(z-zr+r) - 1} \leq \frac{f(r) - f(1)}{r - 1}$$

můžeme převést na

$$\frac{f[r/(z-zr+r)](z-rz+r)}{f(r)} \leq z$$

(4).

Dosazením nerovností (3) a (4) do rovnosti (2) získáme

$$\begin{aligned}
\tilde{A}I_r(e_B | H_f) &\geq \int_0^1 \left(\frac{r-1}{r}\right)^2 z \left(\frac{r-z}{r-1} - z\right) dz \\
&= \int_0^1 \left(\frac{r-1}{r}\right) z(1-z) dz = \frac{1}{6} \frac{r-1}{r} = \tilde{A}I_r(e_B | H_2).
\end{aligned}$$

Dále uvažujme případ $f(0) = 0$. Pak $\tilde{A}I_r(e_B | H_f)$ splňuje

$$\tilde{A}I_r(e_B | H_f) =$$

$$\begin{aligned}
&\frac{1}{a_{r,f}} \int_0^1 \left(\frac{r-1}{r}\right)^2 z \left[f\left(\frac{r}{z}\right) \frac{z}{r} - f\left(\frac{r}{r-zr+z}\right) \frac{r-zr+z}{r} \right] dz + \\
&+ \frac{1}{a_{r,f}} f(0) \frac{r-1}{r} \tilde{A}I_r(e_B | H_2).
\end{aligned}$$

Opět použijeme nerovnosti (3) a (4) (přenásobené $f(r)/r$, abychom se vyhnuli možnému dělení nulou nebo nekladným číslem) a dostaneme

$$\tilde{A}I_r(e_B | H_f) \geq \frac{1}{a_{r,f}}.$$

$$\cdot \left[\left(\frac{f(r)}{r} + f(0) \frac{r-1}{r} \right) \tilde{A}I_r(e_B | H_2) \right]$$

$$\tilde{A}I_r(e_B | H_f) = \frac{r}{f(r) + (r-1)f(0)}.$$

$$\left[\frac{f(r) + f(0)(r-1)}{r} \tilde{A}I_r(e_B | H_2) \right] = \tilde{A}I_r(e_B | H_2).$$

Závěr

Získaný výsledek zahrnuje mocninné entropie H_α pro $\alpha \in (0,2]$ a také další f -entropie, například Emlenův index [13]

$$H_E(p(x)) = \sum_x p(x)e^{-p(x)} - e^{-1},$$

kde

$$f(x) = x \left(e^{-1/x} - e^{-1} \right) \chi_{(0,\infty)}(x),$$

Goodův index [14]

$$H_{G(\alpha,\beta)}(p(x)) = \sum_x p(x)^\alpha (-\ln p(x))^\beta$$

když $\alpha \in [1/2, 1]$, $\beta = 1$, kde

$$f(x) = x^{2-\alpha} (\ln x)^\beta I_{(1,\infty)}(x)$$

a jeho zobecnění, index Sharmy a Taneji [15]

$$H_{ST(s,t)}(p(x)) = (s-t)^{-1} \sum_{i \in M} (p(x)^t - p(x)^s)$$

když $t \in [0,1]$, $s \in (\max\{t, 1-t\}, 3-t]$ (dodefinovaný limitně při $s=t$) kde

$$f(x) = (s-t)^{-1} (x^{2-t} - x^{2-s}) \chi_{[1,\infty)}(x),$$

a Ferreriho index [16]

$$H_{Fe(\alpha)}(p(x)) = (1+\alpha)^{-1} \ln(1+\alpha) -$$

$$-\alpha^{-1} \sum_x (1+\alpha p(x)) \ln(1+\alpha p(x))$$

$$\alpha > 0$$

pro který

$$f(x) = \left[\frac{x(1+\alpha)}{\alpha} \ln(1+\alpha) - \right.$$

$$\left. - \frac{x(x+\alpha)}{\alpha} \ln \frac{x+\alpha}{x} \right] I_{(0,\infty)}(x).$$

Poděkování

Tato práce byla spolufinancována z projektu 1M06014 MŠMT ČR a SVV-2010-265 513. Autor děkuje I. Vajdovi za uvedení do problematiky a děkuje také J. Zvárové a M. Studenému za různé užitečné připomínky, které měli při přípravě práce.

Literatura

- [1] Morales D., Vajda I.: Generalized Information Criteria for Optimal Bayes Decisions. Research report 2274. ÚTIA AV ČR, v.v.i.; 2010.
- [2] Fano R.M.: Transmission of Information: A Statistical Theory of Communications. New York: MIT Press; 1961.
- [3] Kovalevsky V.A.: The problem of character recognition from the point of view of mathematical statistics. Character Readers and Pattern Recognition. New York: Spartan Books; 1968;3-30.
- [4] Vajda I., Vašek K.: Majorization, concave entropies and comparison of experiments. Problems of Control and Information Theory. 1985;14:105-15.
- [5] Maley C.C., Galipeau P.C., Finley J.C., Wongsurawat V.J., Li X., Sanchez C.A., Paulson T.G., Blount P.L., Risques R.A., Rabinovitch P.S., Reid B.J.: Ecological Diversity Measures Predict Cancer Progression In Barrett's Esophagus. Gastroenterology. 2006;131:672-673.
- [6] Chakraborty R.: Mitochondrial DNA Polymorphism Reveals Hidden Heterogeneity within Some Asian Populations. American Journal of Human Genetic. 1990;47:87-94.
- [7] Zvárová J.: On Measures of Statistical Dependence. Časopis pro pěstování matematiky. 1974;99:15-29.

- [8] Kullback S., Leibler R.A.: On Information and Sufficiency. Annals of Mathematical Statistics. 1951;22:7986.
- [9] Shannon C.E.: A mathematical theory of communication. Bell System Technical Journal. 1948;27:379-423,623-56.
- [10] Csizsár I.: Eine informationstheoretische Ungleichung und ihre Anwendung auf den Beweis der Ergodizität von Markoffschen Ketten. Publ. Math. Inst. Hungar. Acad. Sci. 1963;8:85-107.
- [11] Zvárová J., Vajda I.: On Genetic Information, Diversity and Distance. Methods of Inform. in Medicine. 2006;2:173-179.
- [12] Horáček M.: Míry biodiverzity a jejich aplikace. (Measures of biodiversity and their applications.) Master thesis. Prague, Charles university. Supervisor J. Zvárová. 2009.
- [13] Emlen J.M.: Ecology: An Evolutionary Approach. Reading: Addison-Wesley; 1973.
- [14] Good I.J.: The population frequencies of species and the estimation of population parameters. Biometrika. 1953;40:237-64.
- [15] Sharma B.D., Taneja I.J.: Entropy of type (.) and other generalized measures in information theory. Metrika. 1975;22:205-15.
- [16] Ferreri C.: Hypoentropy and related heterogeneity divergence measures. Statistica. 1980;40:55-118.

Kontakt

Mgr. Martin Horáček

Centrum biomedicínské informatiky
Ústav informatiky AV ČR, v.v.i.
Pod Vodárenskou věží 2
182 07 Praha 8
e-mail: horacek@euromise.cz

Metody pro tvorbu časově závislých referenčních intervalů a kvantilových křivek pro hodnocení velikosti plodu

M. Hynek^{1,2}

¹Gennet, Centrum pro fetální medicínu, Praha

²Gynekologicko-porodnické oddělení, Fakultní Thomayerova nemocnice, Praha

Školitelka: Prof. RNDr. Jana Zvárová, DrSc.

Souhrn

Hodnocení velikosti plodu pomocí ultrazvukové biometrie je stěžejní pro správný management těhotenství. K interpretaci těchto měření nám slouží referenční intervaly a kvantilové křivky závislé na gestačním stáří. Pomocí nich identifikujeme extrémní hodnoty, které mohou poukázat na přítomnost patologie. Nejprve jsou zmíněny základní body důležité pro vlastní návrh studií používaných pro konstrukci těchto intervalů. Hlavním cílem práce je podat přehled dostupných statistických metod pro tvorbu referenčních intervalů a kvantilových křivek, jejich metodik, hodnocení shody modelu s daty, jejich výhod a omezení.

Klíčová slova: plod, gestační stáří, ultrazvuk, biometrie, patologie, referenční interval, kvantil, kvantilová křivka, normální rozdělení, normalizující transformace, kvantilová regrese

1. Úvod

Hodnocení velikosti plodu je klíčové pro správný management těhotenství. Téměř bezvýhradně se opírá o ultrazvukové měření embryonálních a fetálních biometrických parametrů (např. temenokostřční délka, obvod hlavičky, délka femuru, apod.), z nichž naprostá většina v průběhu gravidity roste [1]. K jejich správnému posouzení slouží referenční intervaly (RI) a kvantilové křivky, které používáme pro interpretaci získaného biometrického fetálního parametru tak, že ho porovnáváme s velikostí daného parametru v referenční populaci [2].

RI (často nazývané zavádějícím termínem „normální rozmezí“) představují rozmezí mezi dvojicí daných extrémních kvantilů (typicky 5. a 95. percentil pro 90% RI) daného biometrického parametru (závislé veličiny y) pro dané gestační stáří (GS,

nezávislá veličina t). Kvantilová křivka vykresluje hodnoty y odpovídající jednomu nebo více kvantilům pro dané GS v závislosti na GS.

Při hodnocení velikosti je naměřená hodnota y , která leží mimo RI, považována za extrémní a může poukazovat na přítomnost patologie u plodu (např. morfológickou nebo chromozomální vadu [3], intrauterinní růstovou restriktci [4], makrosomii [5]). Další více informativní možností je určit přesný percentil, kterému odpovídá naměřená hodnota, a to ze známé distribuce proměnné y v daném GS. Čím více se získaná hodnota blíží 0. a 100. percentilu, tím extrémnější je dané měření ve srovnání s referenční populací [6].

V literatuře lze nalézt celou řadu postupů pro konstrukci RI a kvantilových křivek. Výběr vhodné metody je pro fetální biometrii klíčový, neboť nepřesně stanovené kvantily mohou vést k chybným klinickým závěrům ohledně vývoje plodu. Cílem této práce je podrobně prozkoumat možné statistické metody pro tvorbu RI a kvantilových křivek pro hodnocení velikosti plodu.

2. Návrh studie a sběr dat

Dříve než-li se budeme věnovat jednotlivým statistickým metodám, zmíníme nejdůležitější body týkající se návrhu studie a sběru dat.

Velikost a růst. Je důležité rozlišovat přesný význam těchto dvou pojmů, které jsou často v praxi zaměňovány. Velikost plodu je dána jedním měřením v daný časový okamžik, zatímco růst plodu představuje dynamický proces [7], [8]. Tyto dvě koncepty v sobě obsahují zcela rozdílné informace. Velikost plodu může být normální navzdory abnormálnímu růstu a naopak, normálně rostoucí plod

může mít abnormální velikost. Přesné rozlišení mezi těmito dvěma pojmy je důležité pro správný výběr typu zamýšlené studie.

Průřezové a longitudinální studie.

Rozdíl mezi průřezovou a longitudinální studií je analogický rozdílu mezi velikostí a růstem plodu [1], [4]. V průřezové studii každý plod přispívá do souboru pouze jedním měřením, zatímco v longitudinální je měření několikrát opakováno v průběhu času. Z tohoto důvodu nám data z průřezových studií dávají informaci pouze o velikosti a mohou být použita pro srovnání velikosti plodu (známého GS) v daný časový okamžik s referenční populací.

Data z longitudinálních studií lze použít pro tvorbu RI pro hodnocení velikosti i růstu plodu. Zpracování těchto dat vyžaduje odlišnou metodiku, která musí zohledňovat odlišnosti dané charakterem longitudinálních studií. Opakovaná měření jednoho plodu jsou vysoce závislá, takže skutečná velikost souboru odpovídá v tomto případě spíše počtu plodů než celkovému počtu měření [1]. Použitá metodika zpracování musí být tomu patřičně přizpůsobena. Pro podrobnější informace týkající se této problematiky odkazujeme na literaturu [9], [10]. V této práci se budeme dále věnovat konstrukci referenčních křivek pro velikost plodu z dat z průřezových studií, které jsou na poli fetální biometrie mnohem častější, pravděpodobně díky snazšímu sběru dat ve srovnání s longitudinálními studiemi.

Výběr vzorku a sběr dat. Výběr adekvátního vzorku je velmi podstatný. Výhodou je, pokud jsou data sbírána čistě pro účely konstrukce RI [1].

Cílem je získat co nejrepresentativnější vzorek populace, nejlépe z vyšetření bez předchozí selekce, neboť např. data použitá z ultrazvukových vyšetření prováděných z nějaké určité klinické indikace mohou výrazně ovlivnit výsledek. Vzhledem k tomu, že výsledek těhotenství je obvykle znám až zpětně, existuje určitá kontroverze ohledně toho, které těhotenství by mělo být považováno za abnormální a vyřazeno z referenčního výběru. Altman a Chitty [1] doporučují použít pro vyřazení pouze informace, které jsou známy v době vyšetření s výjimkou těch gravidit, u kterých je později u plodu zjištěna závažná vrozená vývojová vada nebo onemocnění matky ovlivňující růst plodu, jako např. diabetes mellitus nebo renální onemocnění. Je žádoucí, aby byl každý týden těhotenství zastoupen přibližně stejným počtem měření [1].

Rozsah výběru. Není snadné stanovit, jak rozsáhlý by měl být soubor pro tvorbu kvantilových křivek. Čím větší soubor, tím přesnější budou i odhady percentilů. Zájem klinika je obvykle soustředěn do konců rozdělení. Royston [11] udává, že střední chybu (SE) 100^P% percentilu ($c_{100\alpha}$) lze přibližně vyjádřit jako násobek směrodatné odchylky (SD) pro dané GS vztahem

$$SE(c_{100\alpha}) \approx SD \sqrt{\frac{1 + \frac{1}{2} z_{1-\alpha/2}^2}{n}} \quad (1a),$$

kde $z_{1-\alpha/2}$ je odpovídající hodnota ze standardizovaného normálního rozdělení a n je počet případů [11]. Velikost souboru n můžeme potom přibližně vyjádřit jako

$$n \approx \frac{1 + \frac{1}{2} z_{1-\alpha/2}^2}{\left(\frac{SE(c_{100\alpha})}{SD}\right)^2} \quad (1b).$$

Jestliže například požadujeme, aby pro 90% referenční interval SE byla 5% SD, potom potřebujeme soubor o $n \approx (1 + 0,5 \cdot 1,6452)/0,052 \approx 941$, při SE odpovídající 10% SD $n \approx 235$. Kromě toho, čím extrémnější percentil, tím i méně přesný je jeho odhad. Je tedy zřejmé, že potřebujeme obecně několik set pozorování, abychom pro extrémní percentily získali odhady rozumné přesnosti.

3. Statistické metody

Během posledních dvaceti let se objevila celá řada statistických metod pro tvorbu RI a kvantilových křivek. Mezi základní požadavky kladené na metodu patří tvorba kvantilů, které se mění hladce s průběhem GS a vykazují dobrou shodu s daty [1]. Dříve udávaný požadavek na co nej-jednodušší metodiku [1] se se širokou dostupností výpočetní techniky pomalu vytrácí. Kromě toho je žádoucí, aby model umožnil pro jakékoli v budoucnu získané měření kalkulaci odpovídajícího percentilu a Z skóru [12]. Během posledních let jsou Z skóry stále častěji využívány a jsou zároveň i systémem, který Světová zdravotnická organizace doporučuje pro porovnávání získaných antropometrických měření s referenční populací [13]. V neposlední řadě bylo prokázáno, že Z skóry lze využít i jako vynikající nástroj kontroly kvality, dovolující ultrazvukařům lépe volit referenční grafy vhodné pro jejich praxi [14] nebo kontrolovat kvalitu jejich měření [15].

Obecně lze metody rozdělit na dvě široké skupiny: parametrické (vycházející z předpokladu o specifickém rozdělení) a neparametrické (empirické).

3.1 Metoda průměru a směrodatné odchylky

Tato hojně rozšířená metoda vychází ze základního předpokladu, že v každém GS sledují naměřené hodnoty normální rozdělení s průměrem a SD, které se mění plynule v závislosti na GS. Požadovanou kvantilovou křivku lze potom získat pomocí vztahu

$$c_{100\alpha} = \mu + k \cdot SD \quad (2),$$

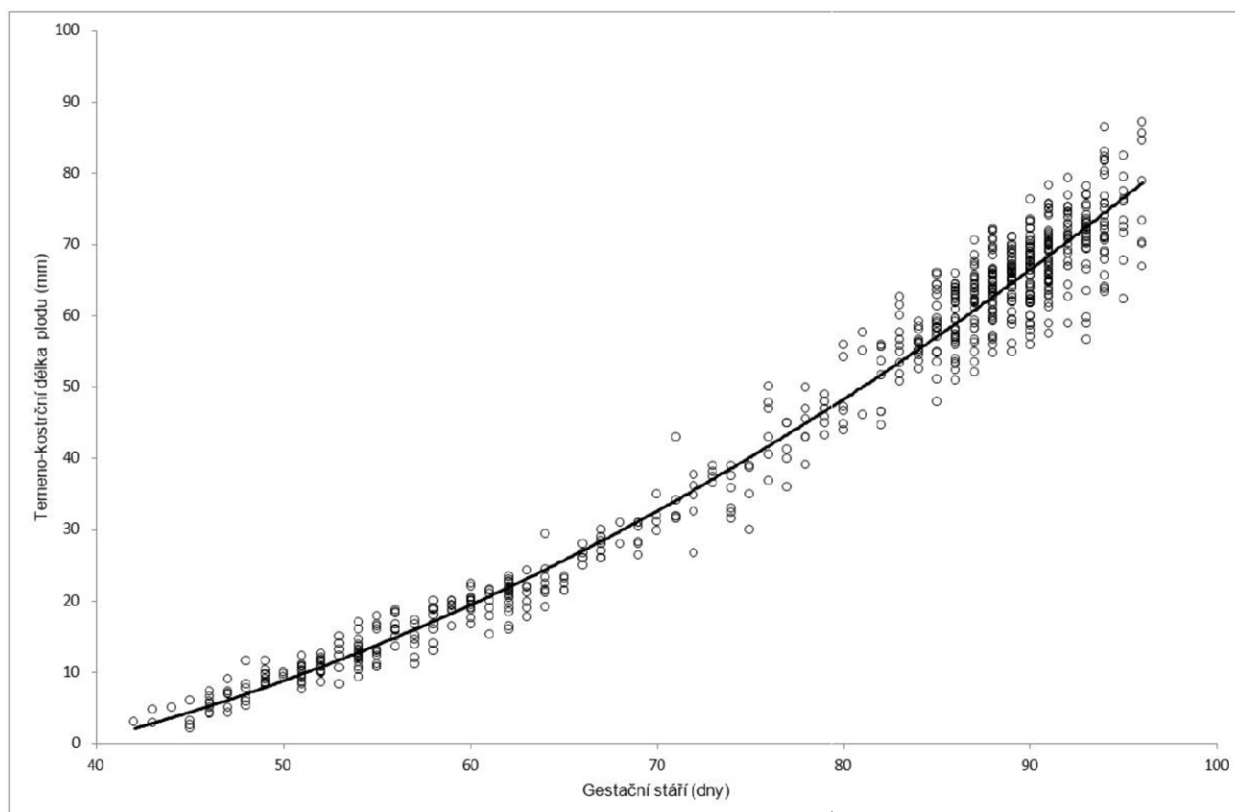
kde k je odpovídající kvantil normálního rozdělení (např. pro 10. a 90. percentil je $k = \pm 1,28$, pro 5. a 95. percentil $\pm 1,645$, apod.) a μ a SD představují průměr a směrodatnou odchylku pro dané GS v referenční populaci. Metoda je založená na modelování křivek pro průměr a SD v závislosti na GS jako polynomicke funkce pomocí metody nejmenších čtverců [1], [4], [11], [16].

Nejprve se z našich dat pokusíme pomocí regresní analýzy vymodelovat adekvátní křivku pro průměr (Obr. 1). Co se týče volby stupně polynomu, Royston [11] a Royston

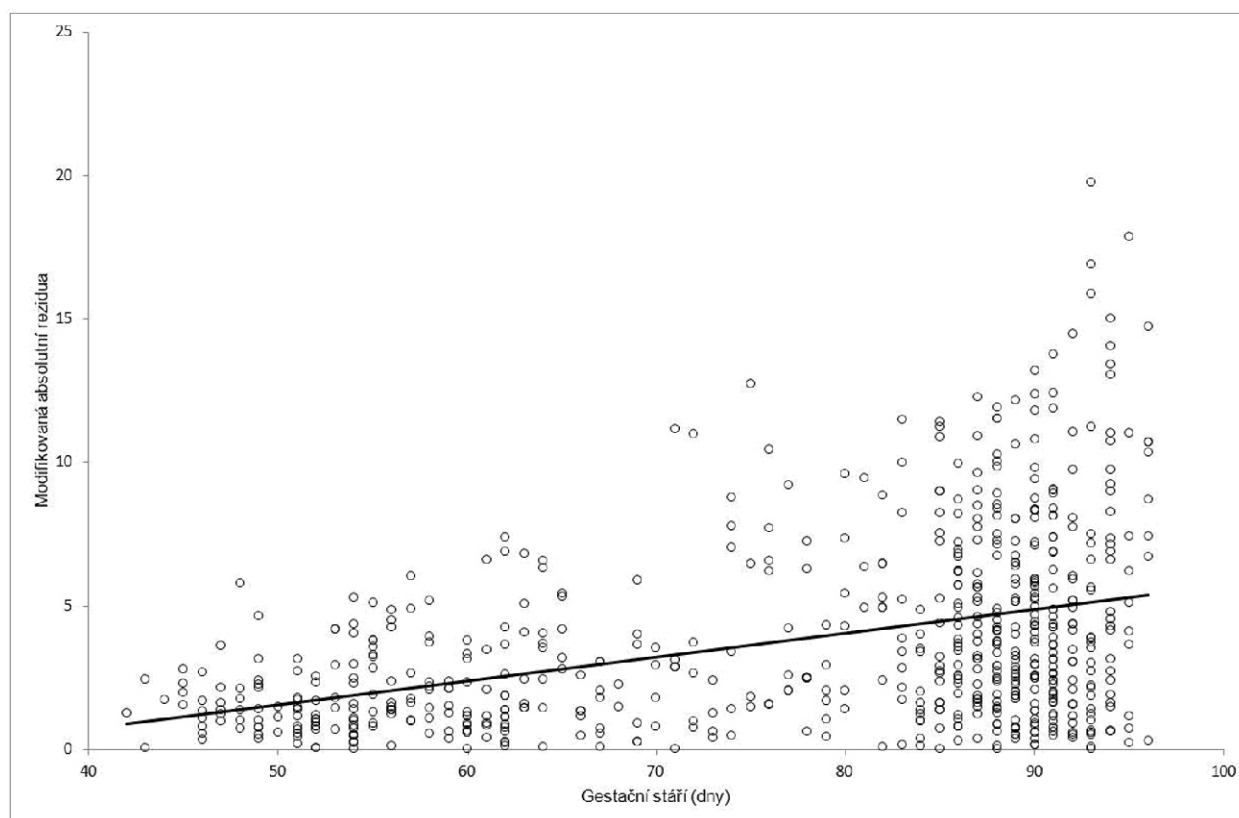
a Wright [4] doporučují primárně začít s kubickým polynommem $(a + b.t + c.t^2 + d.t^3)$. Pokud se koeficient d neliší statisticky významně od nuly (d dělený jeho SE a výsledek porovnán s odpovídající kritickou hodnotou ze Studentova t rozdělení o $n-4$ stupních volnosti), použijeme kvadratický polynom a analogicky opakujeme pro koeficient c . Postup opakujeme, dokud nenarazíme na člen polynomu, který již nelze takto odstranit. Altman a Chitty [7] navrhuji jako alternativu vhodnou pro problematiku velikosti plodu lineárně-kubický model $(a + b.t + c.t^3)$. Prokládání polynomů vyšších stupňů než kubických se nedoporučuje, neboť získané křivky mohou vykazovat nereálné tvary, jako jsou vlnitost nebo prudké změny na začátku a konci rozmezí GS [4]. V případě, že jsou koeficienty pro polynomy čtvrtého a pátého stupně statisticky významné, doporučuje se využít zlomkové polynomy (viz oddíl 3.1.2) [17]. Volba vhodného modelu by neměla vycházet pouze ze statistické významnosti, ale svoji roli by měl mít také estetický vzhled křivky a míra shody s daty.

Další krok zahrnuje posouzení shody získaného polynomu s daty, kdy do grafu vyneseme rezidua (naměřená hodnota minus hodnota očekávaná na základě vytvořeného modelu) v závislosti na GS a sledujeme, zda a jak se mění variabilita v závislosti na čase [7]. Vytvořením normálního grafu reziduí můžeme odhalit odchylku od očekávané normality, např. kladnou nebo zápornou šikmost. V tomto případě bychom měli před vlastní analýzou data transformovat některou z normalizujících transformací. Viz oddíl 3.1.1.

Poté, co získáme vhodný model pro průměr, pokračujeme modelací variability. Mnoho dříve publikovaných prací aplikovalo modely, které nedovolovaly na GS závislou SD. Nicméně, podíváme-li se na oblast fetální biometrie, pak variabilita mezi jednotlivými případy téměř vždy s rostoucí délkou těhotenství roste [7]. Je tedy naprosto nezbytné počítat v modelu s touto časovou závislostí. Postup navržený Altmanem [16] patří mezi nej-používanější.



Obr. 1. Závislost temeno-kostrční délky plodu na gestačním stáří v prvním trimestru těhotenství. Jednotlivými pozorováními je proložena regresní polynomičká křivka modelující průměr (—) pomocí polynomu druhého stupně $y = 0.0126 t^2 - 0,3177 t - 6,7365$. Soubor obsahuje měření 664 plodů s přesně známým gestačním stářím (gravidity po umělém oplodnění).



Obr. 2. Diagram modifikovaných absolutních reziduí v závislosti na gestačním stáří získaných z regresního modelu z Obrázku 1., společně s proloženou lineární regresní křivkou $y = 0,0833 t - 2,6173$.

Základem je předpoklad, že pokud měření všech stárí vykazují normální rozložení, rezidua ze získaného regresního modelu pro průměr by měla mít také normální rozdělení a tím absolutní hodnoty těchto reziduí polonormální rozdělení. Pokud uvažujeme proměnnou X , která sleduje normální rozdělení s průměrem $\mu = 0$ a rozptylem σ^2 , potom absolutní hodnoty $|X|$ mají polonormální rozdělení s průměrem $\sqrt{(2/\pi)}\sigma$ [18]. Průměr absolutních reziduí vynásobený $\sqrt{(\pi/2)}$ je potom odhadem SD reziduí. Tedy, jestliže není SD přijatelně konstantní v závislosti na čase, budou hodnoty předpovězené regresí absolutních reziduí v závislosti na čase a vynásobené $\sqrt{(\pi/2)}$ odhadem SD reziduí pro dané GS a tím i odhadem pro y .

Identický postup navrhl Royston a Wright [4], kdy nejprve absolutní hodnoty reziduí vynásobíme $\sqrt{(\pi/2)}$, čímž vytvoříme tzv. modifikovaná absolutní rezidua (scaled absolute residuals, SARs). Ta posléze regredujeme v závislosti na čase stejným způsobem, jaký byl popsán u modelace průměru, a získaný model použijeme pro odhady SD reziduí (Obr. 2).

Je nepravděpodobné, že bychom k získání uspokojivého modelu potřebovali křivku vyššího stupně než kvadratickou [16]. Pro ověření modelu SD je užitečné vynést $\pm 1,645$ SD do grafu reziduí, neboť bychom očekávali, že 90 % hodnot reziduí bude mezi těmito hranicemi.

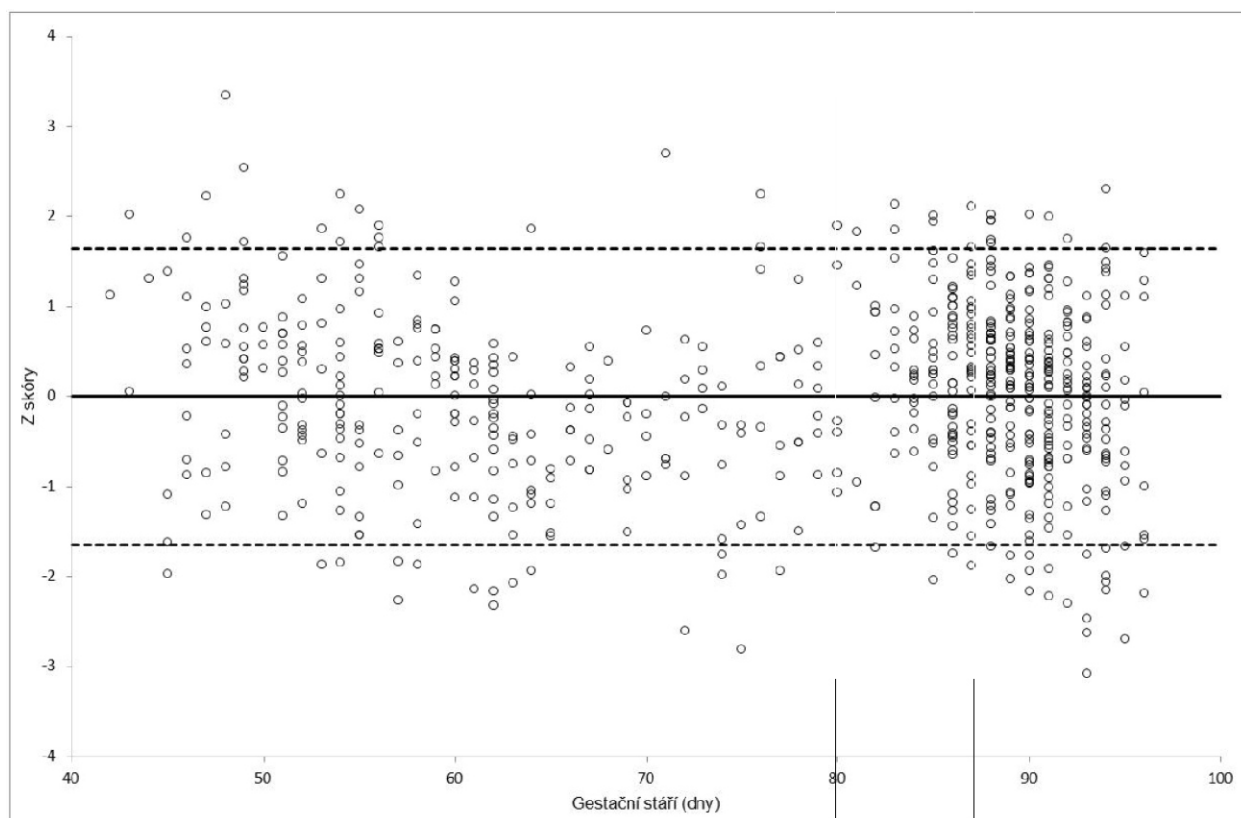
Jestliže se absolutní rezidua nemění významně s GS, je odhad SD shodný se SD původních nemodifikovaných reziduí.

Regresní analýza pro odhad průměru by ve skutečnosti měla brát v úvahu závislost SD na GS. Proto bychom mohli následně znovu vyhladit křivku pro průměr s použitím vážené regrese, kdy jako váhy použijeme převrácenou hodnotu druhé mocniny odhadované SD pro dané GS [19]. Nicméně, efekt na výslednou křivku je téměř vždy nevýrazný [7].

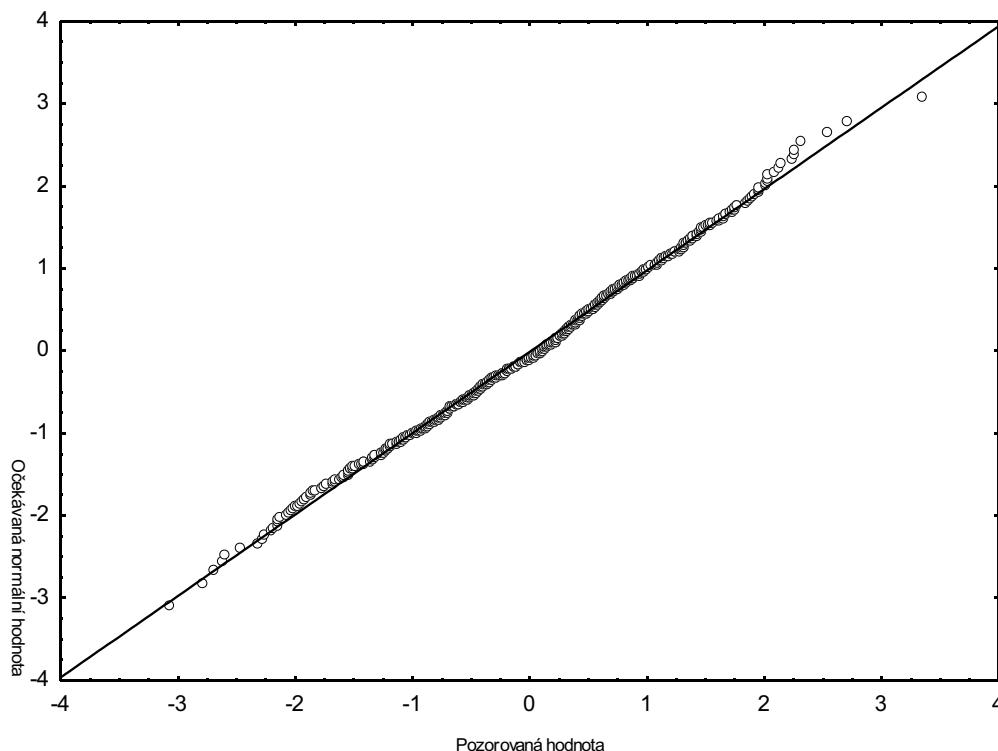
Následující důležitý krok zahrnuje ověření shody modelu s daty. Užitečným nástrojem jsou zde Z skóry, které vypočítáme pro pozorované hodnoty y pomocí vztahu

$$Z \text{ skór} = \frac{y - \mu_{GS}}{SD_{GS}} \quad (3),$$

kde μ_{GS} and SD_{GS} jsou průměr a SD dané našim modelem pro dané GS, v kterém měření proběhlo. Pro hodnocení je doporučováno několik postupů [7]: v grafu Z skórů v závislosti na čase bychom neměli odhalit žádnou systematickou odchylku, hodnoty by měly být náhodně rozmístěny okolo nuly a mezi příslušnými kvantily bychom měli mít očekávaný podíl hodnot (např. 90 % hodnot Z skórů by mělo ležet mezi $Z = \pm 1,645$, 5% nad a 5 % pod) (Obr. 3). Dále lze pro kontrolu, zda hodnoty Z skórů mají přibližně normální rozdělení, použít normální graf (Obr. 4). Body grafu by měly ležet přibližně na přímce. K formálnímu ověření normality lze použít W test (podle Shapira a Wilka) [20] nebo W' test (podle Shapira a Franciové) [21]. Zde je nutno podotknout, že u velkých souborů může již malá odchylka od normality vést k statisticky významné nenormalitě. Je proto rozumné nezamítat normalitu u středně malých p (tj. přibližně $p = 0,01$), jestliže normální graf neukazuje jasnou odchylku od přímky [7].



Obr. 3. Z skóry jednotlivých pozorování s očekávaným 5. a 95. percentilem (-----). 6,02 % pozorování leží pod 5. percentilem, 5,27 % nad 95. percentilem a 88,71 % mezi 5. a 95. percentilem.

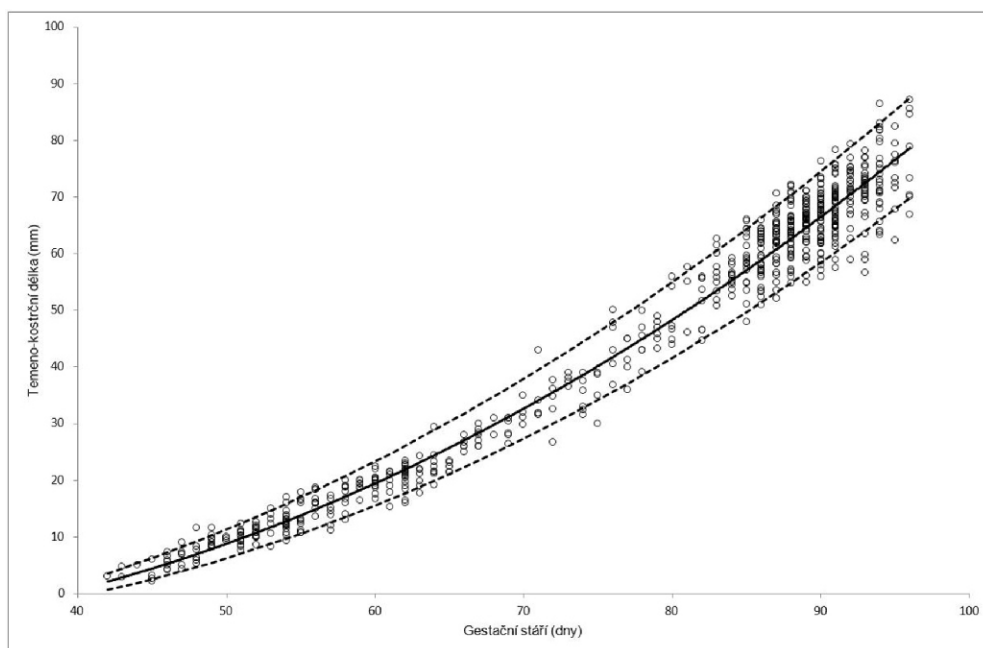


Obr. 4. Normální diagram Z skóre. W test podle Shapira a Wilka $W = 0,997$ ($p = 0,31$).

Nakonec, jestliže jsme s modelem spokojeni, kvantilové křivky pro požadované RI lze vypočítat dosazením vymodelovaného průměru a SD do vztahu (2).

Vykreslením křivek do bodového diagramu společně s pozorováními y slouží k závěrečnému posouzení získaného modelu (Obr. 5).

Z skóre pro jakékoli nové měření lze vypočítat pomocí vztahu (3) a tomu odpovídající percentil získat z inverze distribuční funkce normálního rozdělení $N(0,1)$.



Obr. 5. Získaný model závislosti temeno-kostrční délky plodu na gestačním stáří, vynesena jsou jednotlivá pozorování společně s referenční křivkou pro průměr (—) a křivkou pro 5. a 95. percentil (-----).

3.1.1 Logaritmická transformace

Výše popsaná metoda musí splňovat předpoklad, že v každém GS pocházejí data z populace s normálním rozdělením. Tento požadavek není ale vždy splněn, neboť mnoho fetálních biometrických parametrů vykazuje často nenulovou šikmost (obvykle kladnou šikmost, kdy je pravý konec distribuce protaženější ve srovnání s levým koncem). Logaritmická transformace může problém pomoci odstranit. Mimo to, tato transformace vede ke stabilizaci variability v případě, že SD výrazně roste s GS [11].

V případě, že rezidua z počátečního modelu mají kladnou šikmost, navrhuje Royston [11] provést logaritmickou transformaci původních hodnot y a model znovu vyhladit s použitím $\log(y)$. Jestliže jsou rezidua i z takto nově získaného modelu stále zešikmena, navrhuje použít posunutou logaritmickou transformaci $\log(y + C)$, kdy $C > 0$ pro rezidua zešikmená doleva a $C < 0$ pro rezidua zešikmená doprava. Optimální hodnota C je taková, kdy je hodnota p testů normality reziduí největší.

Po dokončení modelace je důležité nezapomenout na zpětnou transformaci získané křivky užitím antilogaritmu (a event. odečtením C).

3.1.2 Zlomkové polynomy

Konvenční polynomy mají celou řadu dobře známých nevýhod. Polynomy nízkého stupně nabízejí jen omezenou škálu tvarů, a tedy ne vždy lze data proložit adekvátní křivkou a polynomy vyšších stupňů nemusí dobře odpovídat datům na počátku a konci pozorovaného časového rozmezí. Kromě toho postrádají asymptoty a nelze jimi tedy proložit data, u nichž očekáváme mezní chování [22]. Royston a Altman [17] předložili rozšířenou rodinu polynomů, nazývaných zlomkové polynomy (fractional polynomials, FPs), jejichž členy mají exponenty omezené předem stanovenou množinou celočíselných a neceločíselných hodnot. Konvenční polynomy jsou podmnožinou této rodiny. Jestliže je konvenční polynom dán výrazem ve tvaru

$$a + b.t + c.t^2 + d.t^3 + \dots \quad (4),$$

FPs je definován jako

$$a + b.t^{p_1} + c.t^{p_2} + d.t^{p_3} + \dots \quad (5),$$

kde p_1, p_2, \dots jsou vybírány z množiny $\{-2; -1; -0,5; 0; 0,5; 1; 2; 3\}$. Je dáno, že mocnina 0 představuje přirozený logaritmus, tedy t^0 se rovná $\log_e(t)$. Dalším rozšířením jsou „opakující se exponenty“. V tomto případě je druhý člen stejné mocniny vynásoben $\log_e(t)$. Např. FP třetího stupně s exponenty $(0; 2; 2)$ odpovídá výrazu

$$a + b.\log_e(t) + c.t^2 + d.t^2.\log_e(t) \quad (6).$$

FP prvního stupně je výraz ve tvaru $a + b.t^p$. Pro daný soubor dat nalezneme optimální hodnotu p tak, že postupně proložíme data osm samostatných lineárních regresních funkcí s použitím t^2, t^1, \dots, t^p a následně vybereme hodnotu p , pro kterou model vykazuje nejlepší shodu. V případě FP druhého stupně $(a + b.t^{p_1} + c.t^{p_2})$ bude celá procedura výběru nejvhodnějšího modelu čítat již 36 možných kombinací exponentů. Nejlepší model bude potom ten s nejnižší reziduální SD.

Aplikací FP často docílíme lepší shody modelu, a to zároveň s méně členy polynomu, než při použití konvenčních polynomů.

3.2 LMS metoda

LMS metoda, tak jak ji představil Cole [23], [24] a později zdokonalil Cole a Green [25], je velmi flexibilní a široce využitelná semiparametrická metoda schopná tvorby hladkých kvantilových křivek i v případě dat, která mají komplexnější strukturu. Navíc je schopná vypořádat se v čase se měnící šikmostí, kterou klasická logaritmická transformace normalizovat nedokáže. Metoda vychází z předpokladu, že vhodně zvolená mocninná transformace dokáže odstranit šikmost a data normalizovat. Příkladem takovéto normalizující transformace je rodina transformací navržená Boxem a Coxem [26], s transformačním mocninným parametrem λ , který je pro dané GS optimalizován tak, aby kompletně odstranil šikmost.

Rozdělení závisle proměnné y se mění hladce s GS a je kompletně popsáno třemi parametry λ (Box-Coxova mocnina, vztahující se k šikmosti), μ (medián) a σ (variační koeficient), jejichž iniciály (L, M a S) daly metodě jméno.

Předpokládáme, že y má medián μ a že y^λ (nebo $\log_e(y)$ pro $\lambda = 0$) má normální rozdělení. Potom uvažujeme transformovanou proměnnou x

$$x = \begin{cases} \left(\frac{y}{\mu}\right)^\lambda - 1 & \text{pro } \lambda \neq 0 \\ \log\left(\frac{y}{\mu}\right) & \text{pro } \lambda = 0 \end{cases} \quad (7)$$

vycházející z Box-Coxovy mocninné transformace [26]. Tato transformace převádí y na x , které má přibližně normální rozdělení. Pro $\lambda = 1$ odpovídá SDx přesně variačnímu koeficientu (CV) proměnné y , a toto přibližně platí pro λ blízka jedné [25]. Optimální hodnota λ je taková, která minimalizuje SD_x [25].

Označíme-li SD_x (a tedy CV_y) jako σ , potom Z skór proměnné x (a tedy y) je dán

$$Z = \frac{x}{\sigma} = \begin{cases} \frac{\left(\frac{y}{\mu}\right)^\lambda - 1}{\lambda\sigma} & \text{pro } \lambda \neq 0 \\ \frac{\log\left(\frac{y}{\mu}\right)}{\sigma} & \text{pro } \lambda = 0 \end{cases} \quad (8)$$

a předpokládáme, že má standardizované normální rozdělení. Předpokládáme nyní, že rozdělení náhodné veličiny y závisí na čase t (GS) a že λ, μ a σ v daném čase t jsou hodnoty určené z křivek $L(t), M(t)$ a $S(t)$. Pak

$$Z = \begin{cases} \frac{\left(\frac{y}{M(t)}\right)^{L(t)} - 1}{L(t)S(t)} & \text{pro } L(t) \neq 0 \\ \frac{\log\left(\frac{y}{M(t)}\right)}{S(t)} & \text{pro } L(t) = 0 \end{cases} \quad (9).$$

Z rovnice (9) lze odvodit, že kvantil $c_{100\alpha}$ náhodné veličiny y v daném t je dán vztahem

$$c_{100\alpha}(t) = \begin{cases} M(t) \left[1 + L(t) S(t) z_{\alpha} \right]^{1/L(t)} \\ M(t) e^{S(t) z_{\alpha}} \end{cases} \quad \begin{matrix} \text{pro } L(t) \neq 0 \\ \text{pro } L(t) = 0 \end{matrix} \quad (10),$$

kde z_{α} je příslušný α -kvantil standardizovaného normálního rozdělení. Tím je dáno, že jestliže jsou křivky $L(t)$, $M(t)$ a $S(t)$ hladké, jsou hladké i kvantilové křivky. Rovnice (10) nám dovoluje zkonstruovat kvantilovou křivku pro libovolné α a rovnice (9) vyjádřit jakékoli nově získané měření formou Z skóru.

Původní navržená metoda [23], [24] zahrnovala rozdělení dat do co možná nejúžších věkových skupin, ale zároveň s dostatečným počtem případů (pro spolehlivý výsledek alespoň 100), poté odhad λ , μ a σ pro každou skupinu zvlášť pomocí metody maximální věrohodnosti a nakonec proložení získaných hodnot hladkými křivkami $L(t)$, $M(t)$ a $S(t)$, kdy pro vyhlazení lze užít libovolnou metodu (kubické spliny, jádrový odhad, polynomy nebo jednoduše kreslení od oka). Způsob rozdělení do jednotlivých věkových skupin je volitelný a může tedy ovlivnit výslednou podobu křivek.

Cole a Green [25] dodali původní LMS metodě neparametrický charakter tím, že použili pro odhad parametrů metodu maximální penalizované věrohodnosti. Křivky $L(t)$, $M(t)$ a $S(t)$ jsou zde odhadovány maximalizací penalizované věrohodnosti

$$\begin{aligned} \ell - \frac{1}{2} \alpha_{\lambda} \int \{L''(t)\}^2 dt - \\ - \frac{1}{2} \alpha_{\mu} \int \{M''(t)\}^2 dt - \\ - \frac{1}{2} \alpha_{\sigma} \int \{S''(t)\}^2 dt \end{aligned} \quad (11),$$

kde α_{λ} , α_{μ} a α_{σ} jsou vyhlazovací parametry a logaritmická věrohodnostní funkce ℓ je dána vztahem

$$\begin{aligned} \ell = \ell(L, M, S) = \\ = \sum_{i=1}^n \left(L(t_i) \log \frac{y}{M(t_i)} - \log S(t_i) - \frac{1}{2} z_i^2 \right) \end{aligned} \quad (12).$$

Tři integrály zajišťují míru hrubosti pro křivky $L(t)$, $M(t)$ a $S(t)$ a vedou k přirozeným kubickým splinům s uzly v každé rozhodné hodnotě t . Výhodou tohoto postupu je, že subjektivní krok tvorby věkových skupin je odstraněn, s celým souborem dat je zacházeno jako s jednotným celkem a proces vyhlazení křivek je kontrolován přímo pomocí třech vyhlazovacích parametrů. Není nutno určovat žádné věkové hranice jednotlivých skupin a hodnota L , M , a S pro každé t je postupně použita pro výpočet zbývajících dvou parametrů. Jediná subjektivita, která v postupu zůstává, je volba α_{λ} , α_{μ} a α_{σ} nebo alternativně volba ekvivalentních stupňů volnosti (equivalent degrees of freedom, EDF), které jsou funkcí těchto vyhlazovacích parametrů a jsou využitelnějším měřítkem míry vyhlazení. Hodnota EDF pro jednotlivé křivky $L(t)$, $M(t)$ a $S(t)$ odpovídá stupni komplexnosti dané křivky (např. EDF = 1 znamená konstantní průběh, EDF = 2 odpovídá přímce, EDF = 3 přibližně kvadratické křivce a EDF ≥ 4 odpovídá komplikovanějším tvarům). Volba EDF pak musí vyvážit stupeň vyhlazení a zachovat věrnost datům. EDF nízkých hodnot mohou vést k přehlazení, zatímco vysoké EDF k podhlazení křivky. Nevýhodou tedy zůstává, že volba optimálních EDF je víceméně subjektivní. Nicméně, existuje jistá objektivní pomoc pro výběr nejlepší shody modelu založená na skutečnosti, že dvojnásobek změny logaritmické věrohodnosti má přibližně χ^2 rozdělení, kde počet stupňů volnosti je dán změnou EDF [6]. Další možností, která může pomoci s výběrem nejvhodnějších EDF, je tzv. „worm plot“ („červí diagram“), velmi citlivý diagnostický nástroj navržený Buurem a Frederiksem [27]. Worm plot se skládá ze sady detrendovaných normálních grafů reziduí (detrendovaný znamená, že příslušný empirický kvantil je odečten od odpovídajícího očekávaného normálního kvantilu [28]), které jsou

rozděleny do navazujících věkových skupin. Model, který má dobrou shodu, je v diagramu představován přibližně rovnou čarou, kterou autoři přirovnávají k červu. Změna tvaru a lokalizace „červa“ v diagramu je potom charakteristická pro jednotlivé typy neshody (např. v modelaci průměru, rozptylu, šikmosti, špičatosti).

3.3 LMSP metoda

Předpoklad normality po Box-Coxově transformaci může být narušen přítomností špičatosti, kterou transformace nedokáže zohlednit.

Jako zevšeobecnění LMS metody lze chápat LMSP metodu, kterou prezentovali Rigby a Stasinopoulos [29]. Využívá Box-Coxovy mocninné exponenciální distribuce, která může pomoci přítomnou špičatost vyřešit. Distribuce má čtyři parametry označené μ , σ , ν a τ , které lze interpretovat jako vztahující se k poloze (průměr), variabilitě (zhruba CV), šikmosti (transformace k symetrii) a špičatosti (mocninný exponenciální parametr). Výhodou je, že poskytuje flexibilní model zohledňující v souboru přítomnou šikmost i špičatost (jak ploché tak špičaté rozdělení).

Pro odhad parametrů je použito metody maximální penalizované věrohodnosti a následný odhad kvantilů je analogický jako v LMS metodě.

3.4 HRY metoda

Healy, Rabash a Young [30] (odtud název HRY) prezentovali neparametrický přístup využívající Clevelandovu techniku pro vyhlazování bodových diagramů pomocí lokálně vážené lineární regrese [31].

Metoda nepředpokládá žádné konkrétní rozdělení naměřených hodnot pro dané GS a kvantily jsou odhadovány empiricky. Zároveň předpokládá, že jak kvantily samotné, tak intervaly mezi kvantily v daném GS mají hladké chování. To je vynuceno tím, že rozestupy mezi kvantily lze vyjádřit jako polynomy nízkého stupně pro jejich Z skóry.

Hladkých kvantilových křivek je dosaženo dvoustupňovým procesem. V první fázi získáme vybrané kvantily empiricky. Data seřadíme podle vzrůstajícího GS a vybereme okno prvních k měření, kdy k je zlomek celkového počtu pozorování, typicky 5-10 %. Těmito k hodnotami proložíme regresní křivku v závislosti na GS, ze seřazených reziduí určíme požadované kvantily, pokud je třeba, použijeme interpolaci, a získané hodnoty kvantilů vyneseme proti mediánu GS těchto k měření. Tím jsme použili pro výpočet body 1 až k . Postup se opakuje s posunutím okna o jeden případ doprava, tj. s použitím pozorování 2 až $k + 1$, poté 3 až $k + 2$, ... dokud nejsou všechna pozorování použita pro výpočet vybraných kvantilů.

Kvantilová křivka vzniklá v první fázi bude velmi nepravidelná, proto je nutné ji v druhé fázi vyhladit. Hladkou křivku pro i -tý percentil můžeme vyjádřit ve formě polynomu p -tého stupně

$$y_i = a_0 + a_1 t + a_2 t^2 + \dots + a_p t^p \quad (13).$$

Koeficienty a_j pro určité j jsou poté modelovány jako polynomy z_i , kde z_i je Z skóre i -tého percentilu, tedy

$$a_j = b_{j0} + b_{j1} z_i + b_{j2} z_i^2 + \dots + b_{jq_j} z_i^{q_j} \quad (14),$$

kde stupeň q_j polynomu se může lišit pro různé hodnoty j . Tímto způsobem je dosaženo kontroly rozestupů mezi kvantily a zabráněno jejich křížení. Kombinací rovnic (13) a (14) získáváme lineární model, který současně vyhlazuje všechny kvantily odhadnuté v první fázi, a to např. pomocí metody nejmenších čtverců.

Tato metoda dovoluje značnou flexibilitu. Nicméně, v praxi může být nutné velké množství experimentování s volbou nastavitelných parametrů k , p a q_j , než

najdeme optimální hodnoty. Autoři doporučují posuzovat vhodnost navržené sady těchto parametrů pomocí kalkulace bodů, které připadnou mezi přilehlé percentily a porovnávání s očekávanými hodnotami. Pro každé nové pozorování lze teoreticky spočítat odpovídající Z skóre řešením polynommické rovnice, ačkoli tomu v praxi může stupeň polynomu zabránit [32].

Vzhledem k tomu, že některé závisle proměnné pozorované po delší časové období mohou vyžadovat složitější tvar křivky než může jednotlivý polynom nabídnout, Pan a kol. [33] navrhl rozdělit data do navazujících časových skupin, každou skupinu proložit polynomem a k vyhlazení průsečíků křivek použít další člen v polynomu. Alternativní postup prezentovali Goldstein a Pan [34], kdy jsou věkové podskupiny vymezeny na začátku a polynommické křivky jsou aproximovány po částech.

3.5 Neparametrická kvantilová regrese

Další možností pro konstrukci referenčních křivek představuje řada metod využívajících neparametrické kvantilové regrese [35]. Tyto metody dovolují přímý odhad libovolného kvantilu jako hladké funkce nezávisle proměnné (GS) bez nutnosti předpokladu parametrického rozdělení a jsou tedy nenahraditelné v případě, že žádná transformační metoda nedokáže normalizovat data v celém časovém rozmezí. Metoda je dostatečně robustní vůči přítomnosti odlehklých pozorování a navíc je postup plně autonomní, neboť všechny vyhlazovací parametry jsou optimalizovány bez subjektivního zásahu.

Gannoun a kol. [36] publikovali tři přístupy, kdy pro podmíněné kvantilové křivky použili jádrový odhad hustoty, lokálně konstantní jádrový odhad a metody dvou jader. Autoři použili Gaussovo jádro a obdélníkové jádro a také zmiňují doporučení, co se týče odhadu šířky vyhlazovacího okna. Další příklad použití kvantilové regrese prezentoval Wei a kol. [37]. Pro podrobnější popis metod odkazujeme na příslušné práce [36], [37].

Elegantním nástrojem pro posouzení shody modelu s daty je i zde „worm plot“, jak ukázal Buuren [38].

4. Diskuze

Pro tvorbu časově závislých kvantilů existuje několik různých metod, z nichž každá má jak výhody, tak i omezení. Je tedy nepravděpodobné, že bychom našli jednu univerzální metodu, která by byla optimální za všech okolností.

Předností parametrického modelu průměru a směrodatné odchylky je relativní jednoduchost a aplikovatelnost, kdy k jejímu provedení dostačuje většina základních statistických programů. Je nutné zdůraznit, že validita získaných RI závisí významně na dodržení předpokladu normality dat. Modelací SD v závislosti na čase je metoda schopná si poradit s heteroskedasticitou a pomocí logaritmické, popřípadě posunuté logaritmické transformace může někdy odstranit zeshikmenost dat. Nicméně, šikmost měnící se s GS a nenormální špičatost nedokáže snadno překlenout. Výsledné kvantilové křivky a Z skóry lze vyjádřit explicitními rovnicemi. Fakt, že metoda má omezení dané známými nevýhodami polynommických křivek, může být výrazně zlepšen použitím rodiny zlomkových polynomů.

V oblasti fetální biometrie patří tento postup mezi opakovaně využívaný, pravděpodobně i díky tomu, že velká část fetálních ultrazvukových biometrických parametrů dostatečně splňuje předpoklad normality. Nicméně, ne vždy je tato podmínka udržitelná a zde jsou na místě alternativní metody.

LMS metoda využívající penalizované věrohodnosti je vysoce flexibilní a široce využitelnou metodou, která dokáže vytvořit kvantilové křivky bez ohledu na jejich složitost. Umí si poradit se šikmostí kolísající v čase, ale nenormální špičatost může po transformaci zůstat. Velmi užitečným „vedlejším produktem“ metody jsou samotné křivky $L(t)$, $M(t)$ a $S(t)$, které usnadňují podrobnější pohled do vlastní struktury dat. Ačkoli metoda neumožňuje získat vyjádření kvantilových křivek ve formě vzorců, Z skóry pro jednotlivá nová měření lze spočítat poměrně snadno. Na druhé straně má ale metoda několik nevýhod daných procesem vyhlazování.

Patří mezi ně „okrajový efekt“ (nereálné změny v okrajových částech křivky) v případě, že jsou měření méně zastoupena na začátku a konci časového rozmezí a nejednotné vyhlazování v místech, kde jsou měření nesteromerně zastoupena [25].

V posledních letech je LMS metoda stále častěji používána a byla i vybranou metodou pro konstrukci růstových grafů pro Spojené státy americké (the 2000 CDC Growth Charts for the United States [2]). Objevení se specializovaných programů (LMSChartmaker, autoři Cole a Pan [39]) a balíčků pro hlavní statistické programy (**lmsqreg** pro **R**, autor Carey [40]) ještě více tuto metodu zpřístupnily.

Navržená modifikace LMS metody, LMSP metoda je ještě univerzálnější, neboť počítá s přítomností špičatosti v distribuci dat.

Hlavní předností HRY metody je, že a priori nepředpokládá žádné konkrétní rozdělení dat. Metoda je flexibilní a schopna se vypořádat s mnoha růstovými vzorci. Návrhy Pana [33] a Goldsteina a Pana [34] metodu ještě vylepšují. Zápor je, že volba stupně polynomu vyžaduje značnou zkušenost a hodně hledání a není vždy zcela jasné, jak model zlepšit. I když jsou příslušné rovnice dostupné, je kalkulace Z skóru a percentilu pro dané měření jednoduchá pouze v případě velmi základního modelu. Metoda je vnímavá k odlehlým hodnotám a mimo jiné vyžaduje specializovaný software.

Neparametrické metody založené na kvantilové regresi jsou velmi robustní, neobyčejně flexibilní, poskytují mnohem lepší shodu modelu s daty než jiné metody [38] a jejich aplikace je autonomní. Ačkoli je hlubší pohled do problematiky kvantilové regrese mnohem komplikovanější než je tomu u předchozích metod, možnosti, které metoda skýtá, jsou velmi široké. Metoda dokáže odhalit ve struktuře dat charakteristiky, které jiné metody nezachytí a počet oblastí jejího využití prudce roste [38]. Další velkou výhodou této metody je možnost rozšíření konvenčních nepodmíněných modelů závislých pouze na čase na modely začleňující do modelování předchozí růst a další proměnné. Kromě toho počet

implementací pro statistické programy neustále roste (např. balíček **quantreg** pro **R**, autor Koenker [41]).

Nicméně, hlavním nedostatkem zůstává nedostupnost explicitního vztahu pro konverzi nového měření do odpovídajícího percentilu či Z skóru. Vytvořené křivky mohou vykazovat nepravidelnosti na začátku a konci časového rozmezí a jsou obecně esteticky méně přijatelné než křivky získané pomocí ostatních metod [38].

5. Závěr

Cole poznamenal [23], že vyhlazování kvantilových křivek bylo vždy pokládáno tak trochu za černou magii. Předložili jsme několik rozdílných přístupů, jak lze ke konstrukci kvantilových křivek pro hodnocení velikosti plodu přistoupit.

Bližší porozumění, jak jednotlivé metody fungují, může odhalit jejich možnosti a výhody a pomoci při volbě té nejvhodnější pro daný účel. Zároveň znalost jejich nedostatků a omezení může zabránit v chybné interpretaci výsledků a z toho plynoucích chybných klinických závěrů a suboptimální péči.

Mezi jeden z požadavků, které stanovila komise expertů Mezinárodní zdravotnické organizace, patří, aby bylo možné pomocí jednoznačného vzorce převést naměřenou hodnotu do odpovídajícího kvantilu nebo Z skóru [13]. To vylučuje empirické metody jako je HRY metoda a kvantilová regrese. Možnost volby tak v podstatě zůstává kompromisem mezi jednoduchostí a využitelností více omezujících parametrických metod a mezi přizpůsobivější a širě využitelnější, nicméně uživatelsky méně přívětivou, LMS metodou, ev. LMSP metodou. Možnosti aplikace kvantilové regrese na poli fetální medicíny zbývá ještě více prozkoumat.

Poděkování

Tento výstup vznikl v rámci projektu Specifického vysokoškolského výzkumu SVV-2010-265513.

Literatura

- [1] Altman D.G., Chitty L.S.: Design and analysis of studies to derive charts of fetal size. *Ultrasound Obstet Gynecol* 1993; 3: 378-384.
- [2] Kuczmarski R.J., Ogden C.L., Guo S.S.: 2000 CDC growth charts for the United States: Methods and development. Hyattsville: National Center for Health Statistics; 2002.
- [3] Snijders R.J., Sherrod C., Gosden C.M., Nicolaides KH. Fetal growth retardation: associated malformations and chromosomal abnormalities. *Am J Obstet Gynecol* 1993; 168: 547-555.
- [4] Royston P., Wright E.M.: How to construct 'normal ranges' for fetal variables. *Ultrasound Obstet Gynecol* 1998; 11: 30-38.
- [5] Sokol R.J., Chik L., Dombrowski M.P., Zador I.E.: Correctly indentifying the macrosomic fetus: improving ultrasonography-based prediction. *Am J Obstet Gynecol* 2000; 182: 1489-1495.
- [6] Wright E.M., Royston P.: A comparison of statistical methods for age-related reference intervals. *J R Statist Soc A* 1997; 160: 47-69.
- [7] Altman D.G., Chitty L.S.: Charts of fetal size: 1. Methodology. *Br J Obstet Gynecol* 1994; 101: 29-34.
- [8] Salomon L.J.: Early fetal growth: concepts and pitfalls. *Ultrasound Obstet Gynecol* 2010; 35: 385-389.
- [9] Royston P., Altman D.G.: Design and analysis of longitudinal studies of fetal size. *Ultrasound Obstet Gynecol* 1995; 6: 307-312.
- [10] Royston P.: Calculation of unconditional and conditional reference intervals for foetal size and growth from longitudinal studies. *Stat Med* 1995; 14: 1417-1436.
- [11] Royston P.: Constructing time-specific reference ranges. *Stat Med* 1991; 10: 675-690.
- [12] Waterlow J.C., Buzina R., Keller W., Lane J.M., Nichaman M.Z., Tanner J.M.: The presentation and use of height and weight data for comparing nutritional status of groups of children under the age of 10 years. *WHO Bulletin* 1977; 55: 489-498.
- [13] WHO. Physical status: the use and interpretation of anthropometry. Geneva: World Health Organization; 1995.
- [14] Salomon L.J., Bernard J.P., Duyme M., Buvat I., Ville Y.: The impact on choice of reference charts and equations on the assessment of fetal biometry. *Ultrasound Obstet Gynecol* 2005; 25: 559-565.
- [15] Salomon L.J., Bernard J.P., Ville Y.: Analysis of Z-score distribution for the quality control of fetal ultrasound measurements at 20-24 weeks. *Ultrasound Obstet Gynecol* 2005; 26: 750-754.
- [16] Altman D.G.: Construction of age-related reference centiles using absolute residuals. *Stat Med* 1993; 12: 917-924.

- [17] Royston P., Altman D.G.: Regression using fractional polynomials of continuous covariates: parsimonious parametric modelling. *Appl Statist* 1994; 43:429-467.
- [18] Bland J.M.: The half-normal distribution method for measurement error: two case studies. Talk presented in the Statistics and Econometrics seminar series, University of York, February 2005. <http://www-users.york.ac.uk/~mb55/talks/halfnor.pdf> [Accessed 9 October 2010]
- [19] Aitkin M.A.: Modelling variance heterogeneity in normal regression using GLIM. *Appl Statist* 1987; 36: 332-339.
- [20] Shapiro S.S., Wilk M.B.: An analysis of variance test for Normality (complete samples). *Biometrika* 1965; 52: 591-611.
- [21] Shapiro S.S., Francia R.S.: An approximate analysis of variance test for normality. *J Am Statist Assoc* 1972; 67: 215-216.
- [22] McCullagh P., Nelder J.A.: Generalized linear models (2nd ed). London: Chapman&Hall/CRC; 1989.
- [23] Cole T.J.: Fitting smoothed centile curves to reference data. *J R Statist Soc A* 1988; 151: 385-418.
- [24] Cole T.J.: The LMS method for constructing normalized growth standards. *Eur J Clin Nutr* 1990; 44: 45-60.
- [25] Cole T.J., Green P.J.: Smoothing reference centile curves: the LMS method and penalized likelihood. *Stat Med* 1992; 11: 1305-1319.
- [26] Box G.E.P., Cox D.R.: An analysis of transformation. *J R Statist Soc B* 1964; 26: 211-252.
- [27] Buuren S., Fredriks M.: Worm plot: a simple diagnostic device for modeling growth reference curves. *Stat Med* 2001; 20: 1259-1277.
- [28] Wilk M.B., Gnanadesikan R.: Probability plotting methods for the analysis of data. *Biometrika* 1968; 55: 1-17.
- [29] Rigby R.A., Stasinopoulos D.M.: Smooth centile curves for skew and kurtotic data modelled using the Box-Cox power exponential distribution. *Stat Med* 2004; 23: 3053-3076.
- [30] Healy M.J.R., Rabash J., Yang M.: Distribution-free estimation of age-related centiles. *Ann Hum Biol* 1988; 15: 17-22.
- [31] Cleveland W.S.: Robust locally weighted regression and smoothing scatterplots. *J Am Stat Assoc* 1979; 74: 829-836.
- [32] Silverwood R.J., Cole T.J.: Statistical methods for constructing gestational age-related reference intervals and centile charts for fetal size. *Ultrasound Obstet Gynecol* 2007; 29: 6-13.
- [33] Pan H.Q., Goldstein H., Yang Q.: Nonparametric estimation of age-related centiles over wide age ranges. *Ann Hum Biol* 1990; 17: 475-481.
- [34] Goldstein H., Pan H.: Percentile smoothing using piecewise polynomials, with covariates. *Biometrics* 1992; 48: 1057-1068.
- [35] Koenker R.: Quantile regression. Cambridge: Cambridge University Press; 2005.
- [36] Gannoun A., Girard S., Guinot C., Saracco J.: Reference curves based on non-parametric quantile regression. *Stat Med* 2002; 21: 3119-3135.
- [37] Wei Y., Pere A., Koenker R., He X.: Quantile regression methods for reference growth charts. *Stat Med* 2006; 25: 1369/1382.
- [38] Buuren S.: Worm plot to diagnose fit in quantile regression. *Stat Med* 2007; 7: 363-376.
- [39] Cole T.J., Pan H.: LMSChartmaker. <http://homepage.mac.com/tjcole> [Accessed 13 September 2010].
- [40] Carey V.: LMSqreg: An R package for Cole-Green reference centile curves. <http://www.biostat.harvard.edu/~carey> [Accessed 13 September 2010]
- [41] Koenker R.: Quantreg: An R package for quantile regression and related methods. <http://cran.r-project.org> [Accessed 13 September 2010]

Kontakt

MUDr. Martin Hynek

Gennet, Centrum pro fetální medicínu
Kostelní 9

170 00 Praha 7

e-mail: martin.hynek@gmail.cz

Infrastruktura pro ukládání dat a výpočty v biomedicínském výzkumu

T. Kulhánek¹

¹Ústav patologické fyziologie 1. LF UK, Laboratoř biokybernetiky a počítačové podpory výuky
Školitel: Ing. Milan Šárek, Csc.

Souhrn

Infrastruktura jako služba, tedy infrastruktura poskytovaná zákazníkovi formou služby poskytovatele, je jedním z modelu nasazení tzv. cloud computingu, který umožňuje využít datovou a výpočetní kapacitu v cloudu jako množinu fyzických či virtuálních zařízení. Infrastruktura jako služba může být poskytována zvlášť každému výzkumnému projektu a přitom sdílet stejné fyzické kapacity zapojených počítačů a zařízení. V současné době je testováno poskytování infrastruktury jako služby několika projektům v rámci aktivit sdružení CESNET, 1. lékařské fakulty Univerzity Karlovy v Praze (1. LF UK) a Hudební a taneční fakulty Akademie múzických umění v Praze (HAMU). Současný výzkum v oblasti výpočetní fyziologie je náročný na výpočetní kapacitu. Výpočetní úlohy jsou distribuovány počítačům, které jsou poskytovány infrastrukturou. Projekt v oblasti analýzy lidského hlasu je náročný na propustnost počítačové sítě mezi akustickým a video zařízením na lokální straně a analytickou aplikací na straně výkonného serveru. Tento příspěvek popisuje hlavní vlastnosti a výzvy pro infrastrukturu určené pro takovýto typ aplikací. Infrastruktura jako model nasazení v rámci cloud computingu může být vhodná pro mezioborové týmy a pro spolupráci a integraci vysoce specializovaných softwarových aplikací.

Klíčová slova: cloud computing, infrastruktura jako služba, virtualizace, výpočetní fyziologie, identifikace fyziologických systémů, validace fyziologických systémů, protokol vzdálené plochy, grid computing, hlasové pole

1. Úvod

Oblast aplikací, které podporují biomedicínský výzkum v rámci distribuovaných výpočetních systémů, pokrývá několik úloh. Hlavními úlohami jsou výměna, ukládání a vyhledávání dat. Další úlohou je podporovat analýzu dat a umožňovat dlouhotrvající paralelní výpočty. Posledním, ale stejně důležitým úkolem, je poskytování přiměřeného uživatelského prostředí. Důležitou vlastností, kterou musí tyto systémy splňovat, je požadavek na zachování soukromí dat pacienta, a tudíž musí být zaručeno, že k aplikaci a k datům se dostanou jen oprávnění uživatelé. Z toho vyplývá nutnost použít vysokou úroveň zabezpečení přístupu k datům, případně vhodnou anonymizaci dat.

Systémy zaměřující se na výměnu dat mezi různými organizacemi se pokouší optimalizovat datový tok přes počítačovou síť, šifrují data posílaná přes síť, zajišťují jistou míru spolehlivosti doručení a integrují různé nekompatibilní systémy.

Prvním příkladem distribuovaných systémů jsou PACS systémy (Picture Archiving and Communication Systems). DICOM (Digital Imaging and Communications in Medicine) je v současnosti nejpoužívanější standard pro formát a protokol k výměně obrazové medicínské informace v oblasti radiologie. Zabezpečení přenášených dat je ponecháno jiným vrstvám systému. PACS systémy jsou budované s využitím protokolu DICOM a starají se o ukládání a správu medicínských snímků. Většinou se jedná o systémy uzavřené v rámci nemocnice nebo sítě nemocnic pod společným vlastníkem. Časem vyvstal požadavek informačně spojit systémy z různých lokalit, v důsledku čehož se objevili

i problémy vyplývající z nekompatibility různých PACS systémů, které, i když používají DICOM protokol, mají vlastní implementaci pro řízení a správu DICOM snímků [1]. Pro pouhou výměnu DICOM se vyvinuly pomocné systémy s klasickou strukturou centralizovaného úložiště dat a distribuovaného přístupu uživatelů (např. MeDiMed) [2] nebo komunikační centra s možností posílání dat mezi jednotlivými, do jisté míry rovnocennými, nemocničními systémy PACS (např. EPACS [3] nebo ReDiMed).

Jiným příkladem distribuovaného systému pro medicínu byl projekt R-Bay, v jehož rámci byla zkoumána možnost propojovat obecné systémy, včetně výměny DICOM snímků mezi institucemi z různých evropských zemí a zkoumat možnost poskytovat a využívat služby radiologů vzdáleně na mezinárodní úrovni [4].

Dalším příkladem distribuovaných systémů v medicíně jsou systémy využívající výpočetní a/nebo datový grid pro výměnu medicínských obrázků. Projekt Globus Medicus je postaven na gridovém middlewaru Globus Toolkit a poskytuje služby, které prezentují uživateli rozhraní DICOM. Využívání gridového middleware přináší některé benefity jako spolehlivost, bezpečnost a efektivní přenos dat [5].

Příkladem systému náročného na přenos dat je systém budovaný v rámci projektu FONIATR. Jedná se o systém pro podporu foniatrických vyšetření, poskytuje aplikace pro analýzu zvukového záznamu hlasu přes protokol vzdálené plochy (RDP) v operačním systému MS Windows [6].

Příkladem výpočetního systému je systém vyvíjený v rámci projektu IDENTIFIKACE. Výpočetní modely lidské fyziologie se vyvíjí v prostředí MATLAB/Simulink a nově v jazyce Modelica [7]. Současné modely pokrývají komplexně celou známou funkční část lidské fyziologie a znovu používají publikované vztahy a schémata [8]. V rámci projektu se validují modely s konkrétními daty naměřenými na pacientech, tj. identifikují se fyziologické systémy. Některé parametry modelu nemohou být změřeny a proto jsou odhadovány numerickými optimalizačními metodami, které jsou ale náročné na výpočetní čas a odhad parametrů u složitějších modelů trvá několik hodin až dnů. Proto jsou vyvíjeny techniky pro paralelizování výpočtu a distribuci výpočetních úloh ke stolním počítačům v laboratoři a k počítačům v akademických gridových centrech [9].

Pro každý z výše zmíněných systémů se může vybudovat vlastní infrastruktura složená ze serverů propojených počítačovou sítí. Takové budování je však náročné na lidské zdroje, peníze a čas věnovaný na pořízení, instalaci a konfiguraci všech potřebných serverů a zařízení.

2. Metody

Jedním ze způsobů jak zefektivnit práci s budováním výpočetních a datových zdrojů je používat infrastrukturu, která je poskytována v akademickém prostředí formou sdíleného prostředí výpočetních clusterů. Takovou možnost nabízejí výpočetní a datové gridy. Některé projekty mohou mít specifické požadavky na konkrétní prostředí nebo konkrétní verzi softwarové knihovny, které gridové systémy obecně nemusí splňovat. Tento požadavek může řešit virtualizace a cloud computing.

Výpočetní a datový grid

Výpočetní grid je infrastruktura hardwarových a softwarových prostředků, která poskytuje spolehlivý, konzistentní, všepřístupný a levný přístup k špičkovým výpočetním schopnostem [10]. Datový grid může být charakterizován jako integrující architektura pro efektivní a spolehlivé vykonávání dotazů nad daty, které potřebují citlivý management k tera-

bytovým cache, přenosy v řádech gigabitů za sekundu přes globální síť a naplánování přenosu dat s ohledem k superpočítačovým výpočtům, přiměřený odhad výkonu pro vedení výběru replik dat a další pokročilé techniky, které společně maximalizují použití nedostatkových úložišť a počítačových zdrojů [11].

Administrace a údržba gridové infrastruktury vyžaduje poměrně velké úsilí. Toto je typicky zajištěno externím dodavatelem, provozovatelem výpočetních center. Např. národní gridová iniciativa (NGI) v Česku je udržována aktivitou METACENTRUM, což je jedna z aktivit poskytovatele národní výzkumné a vzdělávací sítě sdružení CESNET [12] a spolupracuje s ostatními NGI z okolních zemí v zastřešující iniciativě tzv. European Grid Initiative (EGI).

Systém používající některý z těchto gridů se musí přizpůsobit a používat API poskytované gridovým middlewarem – Globus Toolkit, gLite aj.

Desktop grid

Kdokoliv (obvykle z akademické a výzkumné komunity) se může připojit a využívat grid budovaný národními gridovými iniciativami. Nicméně jen provozovatel NGI udržuje a rozšiřuje tuto infrastrukturu a gridový middleware.

Rozdílnou filozofií budování gridu je koncepce, kdy kdokoliv může rozšířit gridovou infrastrukturu tím, že do něj připojí vlastní počítač a rozhodne které gridové aplikace mohou jeho počítač využívat k výpočtům. Nejvíce známý případ takového gridového systému je projekt SETI@home [13]. K SETI@home se může připojit kdokoliv, kdo si stáhne nejprve malý obecný klientský program a připojí se již k tomuto konkrétnímu projektu a klientský program se již stará o plánování, stahuje a vykonává výpočetní úkoly na pozadí. Takové systémy jsou obecně označovány jako dobrovolné gridové systémy nebo desktop gridy. Obecný systém pro desktop grid BOINC je používán pro budování přizpůsobené serverové strany a klientské aplikace k formování přizpůsobeného vlastního desktopu gridu [14], [15].

Cloud computing

Cloud computing je konstrukt, který umožní přístup k aplikaci, jenž sídlí v jiné lokalitě než je uživatelův počítač [16]. Cloud computing poskytuje jako službu buď software nebo platformu nebo celou infrastrukturu. Infrastruktura jako služba (Infrastructure as a Service IaaS) se skládá z datových center, výpočetních zdrojů a sítě. Podobně jako v grid computingu, uživatel infrastruktury typicky není vlastníkem cloudu a nestará se o údržbu fyzického hardwaru. Nicméně může mít jistá administrátorská práva k poskytovaným virtuálním zdrojům a může předpokládat, že nikdo jiný koho neurčí, k těmto zdrojům přistupovat nemůže, n rozdíl od grid computingu, kde se zdroje včetně uživatelského prostředí sdílí.

Ke vzdálené aplikaci se může přistupovat podle koncepce tenkého klienta (thin client) a může být realizován např. přes web (protokol HTTP) nebo přes vzdálenou plochu (protokoly RDP, VNC). Oproti aplikaci nasazené na lokálním počítači, může nasazení v cloud computingu vyžadovat jisté úpravy komunikačních protokolů pro přenos netypických dat z klientského počítače.

Virtualizace

Virtualizace je technologie, která izoluje softwarovou vrstvu, na které běží operační systém, od vrstvy fyzického hardwaru. Tím umožňuje provoz jednoho, či více tzv. virtuálních strojů, které sdílí fyzickou paměť, procesory, síťové připojení a vstupně výstupní zařízení. Virtuální stroj může být plně nebo částečně (paravirtualizace) oddělen od fyzické vrstvy hardwaru, v obou případech mohou různé virtuální stroje na nichž běží různé operační systémy (Windows, Linux) pracovat společně a souběžně na jednom fyzickém stroji. Virtualizační techniky přinášejí i jisté režijní náklady spojené s emulací prostředí a při překladu instrukcí izolované aplikace k nižším vrstvám systému. Nicméně bylo ukázáno, že open-source paravirtualizační systém XEN nezpůsobuje výraznou penalizaci ve výkonu virtuálního stroje oproti nevirtualizované konfiguraci [17].

Virtualizace ve svém důsledku je klíčovou technologií, která umožňuje cloud computing a provoz různých izolovaných systémů na sdíleném hardwaru.

Virtuální infrastruktura

Virtuální organizace je skupina uživatelů, která se domluvila na sdílení stejných zdrojů. Virtuální infrastruktura patřící virtuální organizaci se skládá z několika virtuálních strojů propojených navzájem sítí, která může být také virtuální a přístupná pouze uživatelům ve virtuální organizaci. Na Obr. 1 je znázorněn příklad několika virtuálních organizací a jejich infrastruktury. V pravé polovině je schématicky znázorněno fyzické propojení různých organizací (nemocnic, výzkumných ústavů) přes akademickou síť či Internet, uzly jsou fyzické zdroje (servery) a hrany síťové propojení mezi nimi. Každý obláček znázorňuje jednu samostatnou virtuální infrastrukturu. V levé části je znázorněn fyzický server, na němž běží více virtuálních strojů, každý náležející jiné virtuální infrastruktuře.

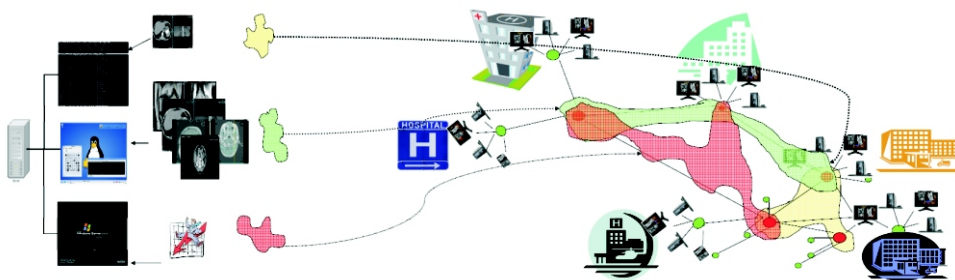
3. Výsledky

Pilotní fyzická infrastruktura určená pro medicínské aplikace byla ustavena v rámci aktivity sdružení CESNET, Podpora aplikací, v několika lokalitách v Praze. Využívá open-source virtualizaci systému XEN a nástroje operačního systému Linux pro konfiguraci virtuálních strojů a virtuální infrastruktury jako služby.

První virtuální infrastrukturou byla množina virtuálních strojů s instancemi gridového systému Globus MEDICUS. Bylo ukázáno, že takovýto gridový systém může být snadno integrován se stávajícími medicínskými systémy používajícími DICOM formát [18].

Kromě systému na výměnu medicínských snímků byla do pilotní infrastruktury nasazena i aplikace z projektu FONIATR. K této aplikaci bylo nutné rozšířit standardní protokol RDP o možnost přenosu audio signálu z mikrofону lokálního počítače ke vzdálené aplikaci [6]. Tento systém v současné době využívá jeden virtuální stroj, viz Obr. 2.

Dále je v této infrastruktuře provozován systém pro identifikaci fyziologických



Obr. 1. Obecné schéma virtuálních infrastruktur.

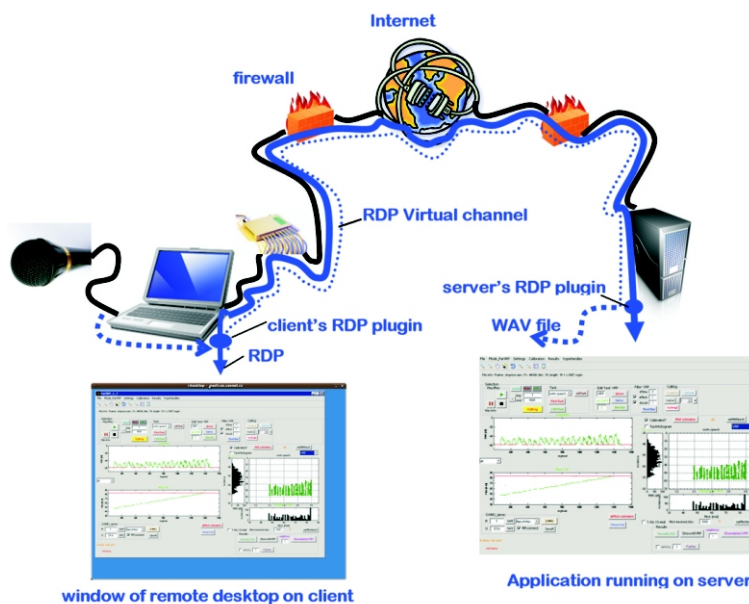
systémů, který poskytuje webovou službu rozdělující výpočetní úlohy stolním počítačům zapojených přes desktop grid systém BOINC a SZTAKI Desktop Grid API [9]. Schéma na Obr. 3 znázorňuje architekturu systému. Server je v provozu jako samostatný virtuální stroj a obsahuje webovou službu. Několik BOINC workerů je také v provozu jako samostatné virtuální stroje na méně využívaných strojích pilotní infrastruktury a navíc je do tohoto desktop grid systému připojeno několik počítačů v laboratoři a učebně na 1. LF UK. Další počítače se mohou snadno do tohoto systému připojit. V současnosti je zkoumána možnost využití výpočetní kapacity NGI MetaCentrum a možnost využít GPU computing.

Všechny fyzické servery pilotní infrastruktury mají virtuální prostředí postavené na systému XEN a sdílejí síťové IP adresy. Každý z virtuálních serverů má vlastní diskový oddíl v rámci Logical Volume Managmentu (LVM) na fyzickém serveru. Virtuální stroje jsou administrovány nástrojem virt-manager a síťové prostředí je na každém fyzickém stroji konfigurováno nástrojem iptables.

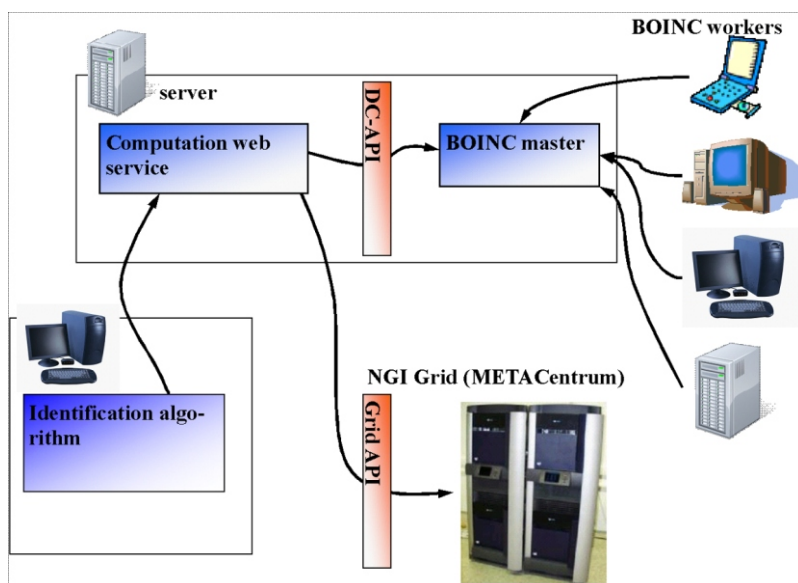
4. Diskuse

Pilotní infrastrukturu lze charakterizovat jako privátní cloud, tj. cloud, který je přístupný jen omezené komunitě uživatelů z oblasti biomedicínského výzkumu.

Virtuální stroje sdílí fyzické síťové připojení přes protokol Ipv4. Kvůli nedostatku unikátních IPv4 adres je konfigurace síťových služeb (webový server, RDP



Obr. 2. Schéma systému pro analýzu zvuku a vzdálené nahrávání přes protokol RDP.



Obr. 3. Schéma výpočetní infrastruktury pro intifikaci fyziologických systémů.

protokol) realizována překladem adres a mapováním portů. Přechodem síťových prvků propojující fyzické servery k IPv6 se otevře možnost přidělování IP adresy virtuálním strojům a konfigurace mapování portů a překladu adres už nebude nutná.

Pro administraci cloudu v pilotní infrastruktuře se nepoužívají zatím žádné speciální nástroje, neboť projektů v cloudu je zatím relativně málo. Nicméně existují produkty (např. Eucalyptus, OpenNebula, VMWare vSphere), které poskytují sadu nástrojů pro automatizaci správy privátního cloudu, včetně konfigurace privátní sítě, migrace virtuálního stroje na jiný fyzický server aj. Nasazení těchto nástrojů bude nezbytné v budoucnu, kdy se očekává rozšiřování fyzické kapacity, případně připojení stávajících zdrojů do větších infrastruktur, které se plánují budovat v rámci akademického prostředí v České republice.

Zmíněný pilotní cloud využívá počítače umístěné fyzicky v různých geograficky vzdálených lokalitách propojené akademickou sítí. V rámci superpočítačových center univerzit se v poslední době začínají budovat také cloudy, které jsou nasazeny na výkonných serverech a clusterech a propojeny lokální sítí, případně rychlými spojeními typu Infini-

Band, které jsou o několik řádů rychlejší než klasické optické trasy používané na delší vzdálenosti.

Důležitá otázka je, pro jaký typ aplikací je vhodné prostředí poskytované cloudem provozovaným na geograficky vzdálených fyzických zdrojích (serverech a clusterech) oproti cloudům v prostředí superpočítačových center. Cloud v superpočítačovém centru je velmi vhodný na vysoce paralelní výpočty vyžadující rychlou komunikaci mezi paralelními výpočetními jednotkami.

Cloud využívající zdroje z různých lokalit může nabídnout volnou kapacitu v době, kdy vlastník svůj fyzický zdroj zrovna nevyužívá a nabízí ho někomu jinému.

5. Závěr

Na fyzické infrastruktuře je možné provozovat privátní cloud a uživatelům poskytovat virtuální infrastrukturu, na které mohou spouštět své vlastní aplikace a systémy. Poskytování infrastruktury jako služby může otevřít přístup k distribuovaným systémům širšímu okruhu uživatelů, kterým dříve bránila příliš složitá administrace, zdoluhavý proces pořízení a instalace vlastních výpočetních zdrojů.

Cloud propojující fyzické počítače v různých lokalitách může být vhodným doplňkem ke cloudům v superpočítačových centrech.

Poděkování

Tato práce vznikla za podpory grantu SVV-2010-265 513.

Literatura

- [1] Kulháněk T., Šárek M.: Virtualizace a integrace v gridovém PACS systému, MEFANET 2008, Brno, ISBN 978-80-7392-065-4
- [2] Dostal, O., Javorník, M., and Ventruba, P. "Collaborative environment supporting research and education in the area of medical image information". International Journal of Computer Assisted Radiology and Surgery. Springer Science+ Business Media. 1. 98. 2006.
- [3] E-PACS, DICOM komunikace mezi zdravotnickými zařízeními, 2010, czech language, web: <http://www.epacs.cz/>
- [4] Pedersen C. D., Rasmussen J., Beck C.: R-Bay; Creating an eMarketplace for the transfer of imaging related eHealth services in Europe. Final project report, version 2, web: <http://www.r-bay.org>
- [5] Erberich S.G., Silverstein J.C., Chervenak, A., Schuler, R., Nelson, M.D., and Kesselman, C.: "Globus MEDICUS- Federation of DICOM Medical Imaging Devices into Healthcare Grids". Studies in Health Technology and Informatics. IOS Press; 1999. 126. 269. 2007.
- [6] Kulháněk T., Frič M., Šárek M.: Remote Analysis of Human Voice Lossless Sound Recording Redirection, Proceedings of Biosignal 2010: Analysis of Biomedical Signals and Images, ISSN 1211-412X, ISBN 978-80-214-4106-4
- [7] Stodulka, P., Privitzer, P., Kofránek, J., Tribula, M., and Vacek, O.: Development of Web Accesible Medical Educational Simulators". Proceedings of EUROSIM 2007, Ljubljana. 2. 16.
- [8] Abram S.R., Hodnett B. L., Summers R.L., Coleman T.G., Hester R.L.: Quantitative circulatory physiology: an integrative mathematical model of human physiology for medical education. Adv. Physiol. Educ., 31, 202-210, (2007).
- [9] Kulháněk T., Šilar J., Mateják M., Privitzer P., Kofránek J., Tribula M.: Distributed computation and parameter estimation in identification of physiological systems, VPH conference 2010, Brussels
- [10] Foster I., Kesselman C. "Computational grids", Vector and Parallel Processing, VECPAR 2000, 2001 Springer

- [11] Chervenak A., Foster I., Kesselman C., Salisbury C., Tuecke S.: The Data Grid: Towards an Architecture for the Distributed Management and Analysis of Large Scientific Datasets. Journal of Network and Computer Applications, Elsevier, 2000
- [12] Šustr Z., Šitera J., Mulač M., Ruda M., Antoš D., Hejtmánek L., Holub P., Salvat Z., Matyska L.: MetaCentrum, the Czech Virtualized NGI, 2009, web: <http://egee.cesnet.cz>
- [13] Anderson D. P., Cobb J., Korpela E., Lebofsky M., and Werthimer, D. 2002. SETI@home: an experiment in public-resource computing. Commun. ACM 45, 11 (Nov. 2002), 56-61.
- [14] Anderson D. P. 2004. BOINC: A System for Public-Resource Computing and Storage. In Proceedings of the 5th IEEE/ACM international Workshop on Grid Computing (November 08 - 08, 2004). International Conference on Grid Computing. IEEE Computer Society, Washington, DC, 4-10.
- [15] Balaton Z., Gombás G., Kacsuk P., Kornafeld A., Kovács J., Marosi C.A., Vida G., Podhorszki N., Kiss T.: SZTAKI Desktop Grid: a Modular and Scalable Way of Building Large Computing Grids. Proceedings of the 21st International Parallel and Distributed Processing Symposium (IPDPS 2007), 26-30 March 2007, Long Beach, California, USA
- [16] Velte T., Velte A., and Elsenpeter R.: Cloud Computing: A Practical Approach". McGraw-Hill Osborne Media. 2009.
- [17] Youseff L., Wolski R., Gorda B., Krintz, C. Paravirtualization for HPC Systems. Frontiers of High Performance Computing and Networking "ISPA 2006 Workshops. 474486. 2006.
- [18] Kulháněk T., Šárek M.: Processing of medical images in virtual distributed environment. ACM. 13. 2009.

Kontakt**Mgr. Tomáš Kulháněk**

Ústav patologické fyziologie 1. LF UK,
Laboratoř biokybernetiky a počítačové
podpory výuky
U Nemocnice 5
128 53 Praha 2
e-mail: tomaton@centrum.cz

Jazyk českých lékařských zpráv a klasifikační systémy v medicíně

P. Přechová¹

¹Centrum biomedicínské informatiky, Ústav informatiky AV ČR, v.v.i.

Školitelka: Prof. RNDr. Jana Zvárová, DrSc.

Souhrn

Cílem příspěvku je porovnat české lékařské zprávy psané volným textem a pomocí softwarové aplikace; analyzovat využitelnost mezinárodních klasifikačních systémů v českém zdravotnickém prostředí. Analýza lékařských zpráv byla založena na attributech Minimálního datového modelu pro kardiologii (MDMK). Použili jsme lékařské zprávy psané volným textem a lékařské zprávy ze softwarové aplikace ADAMEK, kde jsou data ukládána strukturovaně. Pro naši práci jsme využili SNOMED CT a MKN-10. Zaměřili jsme se na jazyk česky psaných lékařských zpráv a na aplikaci výše zmíněných mezinárodních klasifikačních systémů v MDMK. Porovnali jsme, jak dobře jsou atributy MDMK zaznamenány v textových lékařských zprávách a v lékařských zprávách zaznamenávaných strukturovaně pomocí softwarové aplikace ADAMEK. Provedli jsme jazykovou analýzu českých textových lékařských zpráv. Dospěli jsme tedy k závěru, že zapisování lékařských zpráv formou volného textu je velice nehomogenní a není standardizováno. Standardizovaná terminologie by přinesla výhody lékařům, pacientům, administrátorům, softwarovým vývojářům a plátcům. Pomohla by poskytovatelům zdravotnické péče tím, že by poskytovala kompletní a snadno dostupné informace, které náleží k procesu zdravotnictví a to by vyústilo v lepší péči o pacienty. Použití mezinárodních klasifikačních systémů je nezbytným prvním krokem, který umožní interoperabilitu heterogenních elektronických zdravotních záznamů.

Klíčová slova: terminologie, synonyma, klasifikační systémy, tezaurus, nomenklatura, elektronický zdravotní záznam, interoperabilita, sémantická interoperabilita, kardiologie, ateroskleróza

1. Úvod

Vymezení, pojmenování a třídění lékařských pojmů není optimální. Dokladem je skutečnost, že pro jeden pojem existuje často více než deset synonym.

Se zaváděním výpočetní techniky v lékařství se tento problém prohloubil, neboť její využívání předpokládá větší jednoznačnost zadávání dat, vymezení pojmů a jejich přesné pojmenování.

Obecně je velmi výhodné využívat v odborné terminologii pro jeden pojem vždy pouze jediný výraz. Synonyma lze sice počítač naučit, zvětšují však rozsah slovníku databáze i počet nezbytných operací, což prodlužuje komunikaci. Synonymie v odborné terminologii vede při sdělování informací navíc k nepřesnostem a nedorozuměním. V současné lékařské terminologii se lze setkat s řadou synonym pro jediné onemocnění. Z tohoto důvodu začaly vznikat kódovací systémy, které rychle poskytnou kód pro libovolný biomedicínský poznatek.

V současné době dochází k velkému rozvoji elektronických zdravotních záznamů. Existuje obecná shoda, že elektronické lékařské zprávy mají potenciál zlepšit kvalitu lékařské péče [1]. V konceptu elektronického zdravotního záznamu je pacient chápán jako aktivní partner, který má přístup ke svým zdravotním datům, může je přidávat a spravovat [2].

Bezpečná a vhodná výměna klinických informací mezi různými elektronickými zdravotními záznamy je nezbytná k zajištění kontinuity péče o pacienty a to v různých časech, na různých místech a u různých poskytovatelů zdravotní péče. Mapování atributů elektronických zdravotních záznamů na mezinárodní klasifikační systémy je tedy důležitým krokem

pro sémantickou interoperabilitu mezi těmito různými systémy. Elektronické zdravotní záznamy a sémantická interoperabilita jsou velice aktuálními problémy a jsou diskutovány v mnoha článcích [3], [4], [5], [6]. Sémantická interoperabilita založená na českém jazyce byla zkoumána [7], [8], [9], [10], [11].

Klinická data z elektronických zdravotních záznamů tradičně obsahovala malé množství dat z fixních polí (často získané ze seznamu možných odpovědí) a většího množství volného textu [12]. V tomto článku se zaměřím na oba dva druhy těchto zpráv.

2. Kódovací a klasifikační systémy

Kódovací systémy omezují variabilitu vyjadřování. Lze používat pouze schválené termíny a jejich spojení a to podle přesně stanovených pravidel. Obvykle jsou namísto schválených termínů používány formální kódy. V mnoha případech je užitečné, když kódovací systém rovněž ukazuje neschválené termíny, které jsou užívány jako synonyma pro schválené termíny.

Ukažme si nyní několik nejrozšířenějších mezinárodních klasifikačních systémů.

2.1 MKN – Mezinárodní klasifikace nemocí a přidružených zdravotních problémů

Základy Mezinárodní klasifikace nemocí (ICD – International Classification of Diseases and Related Health Problems) položil William Farr v roce 1855. V roce 1948 ji přejala Světová zdravotnická organizace WHO. V této době šlo již o 6. verzi. V současné době se používá 10. revize MKN, která platí od roku 1994 a obsahuje 22 kapitol. MKN se stala mezinárodním standardem pro klasifikaci

diagnóz a pro mnohé epidemiologické a managementské potřeby ve zdravotnictví. Mezi ně patří obecná situace o zdraví v různých skupinách populace a monitorování incidence a prevalence nemocí a dalších zdravotních problémů ve vztahu k dalším proměnným. Používá se ke klasifikaci nemocí a dalších zdravotních problémů, které jsou zaznamenány v mnoha druzích zdravotních záznamů, včetně úmrtních listů a nemocničních záznamů. [13]. MKN je k dispozici v šesti oficiálních jazycích WHO (arabštině, čínštině, angličtině, francouzštině, ruštině a španělštině) a v dalších 36 jazycích, včetně češtiny.

2.2 SNOMED CT

SNOMED Clinical Terms vzniknul spojením dvou terminologií: SNOMED RT a Clinical Terms Version 3 (Read Codes CTV3). SNOMED RT znamená **S**ystematized **N**omenclature of **M**edicine **R**eference **T**erminology, kterou vytvořila College of American Pathologists. Slouží jako společná referenční terminologie pro shromažďování a získávání zdravotnických dat zaznamenaných organizacemi nebo jednotlivci. Clinical Terms Version 3 (Read Codes CTV3) vznikla v United Kingdom's National Health Service v roce 1980 jako mechanismus pro ukládání strukturovaných informací o primární péči ve Velké Británii.

V roce 1999 se tyto dvě terminologie spojily a vznikl tak SNOMED CT, což je vysoce komplexní terminologie. V současné době existuje americká, britská, španělská a německá verze SNOMED CT.

2.3 MeSH

Medical **S**ubject **H**eadings je slovník kontrolovaný Národní lékařskou knihovnou (NLM) v USA. Tvoří ho skupina pojmů, které hierarchicky pojmenovávají klíčová slova a tato hierarchie napomáhá při vyhledávání na různých úrovních specifičnosti. Klíčová slova jsou uspořádána jak abecedně tak hierarchicky. NLM využívá MeSH k indexování článků z předních světových biomedicínských časopisů pro databázi MEDLINE/PubMED. MeSH se využívá také pro databázi katalogizující knihy, dokumenty a audiovizuální materiály. Každý bibliografický odkaz je spojován se skupinou termínů v klasifikačním systému MeSH. Vyhle-

dávací dotazy používají také slovní zásobu z MeSH, aby našly články na požadované téma. Specialisté, kteří MeSH slovník vytvářejí, ho průběžně aktualizují a kontrolují. Sbírají nové pojmy, které se začínají objevovat ve vědecké literatuře nebo ve vznikajících oblastech výzkumu, definují tyto pojmy v rámci obsahu existujícího slovníku a doporučují jejich přidání do slovníku MeSH.

Existuje i český překlad MeSH.

2.4 LOINC

Klasifikační systém Logical Observations Identifiers Names and Codes (LOINC) [18] je klinickou terminologií důležitou pro laboratorní testy a laboratorní výsledky. V roce 1999 byl LOINC přijat organizací HL7 jako preferované kódování pro názvy laboratorních testů a klinických pozorování.

2.5 ICD-O

Klasifikační systém ICD-O [19] je rozšířením Mezinárodní klasifikace nemocí (ICD) pro kódování onkologie. Jedná se o čtyřdimenzionální systém. Dimenze jsou určeny pro třídění morfologických typů nádorů. V současné době existuje její třetí verze.

2.6 TNM

TNM [20] klasifikace je klinická klasifikace maligních nádorů, která se využívá pro účely srovnávání terapeutických studií. Vychází z poznatku, že pro prognózu onemocnění je zvláště důležitá lokalizace a šíření tumoru.

2.7 Další klasifikační systémy

V současné době existuje v medicíně více než 100 různých klasifikačních systémů. Mezi ně patří i *AI/RHEUM*; *Alternative Billing Concepts (ABC)*; *Alcohol and Other Drug Thesaurus (AOD)*; *Beth Israel Vocabulary*; *Canonical Clinical Problem Statement System (CCPSS)*; *Current Dental Terminology (CDT)*; *DSM (Diagnostic and Statistical Manual of Mental Disorder)*; *Medical Entities Dictionary (MED)*; *Current Procedural Terminology (CPT)*; *International Classification of Primary Care (ICPC)*; *McMaster University Epidemiology Terms*; *CRISP Thesaurus*; *Coding Symbols for a Thesaurus of Adverse Reaction Terms (COSTART)*; *Diseases Database*;

DXplain; *Gene Ontology (GO)*; *Healthcare Common Procedure Coding System (HCPCS)*; *Home Health Care Classification (HHCC)*; *Health Level Seven Vocabulary (HL7)*; *Master Drug Data Base (MDDDB)*; *Medical Dictionary for Regulatory Activities Terminology (MedDRA)*; *Multum MediSource Lexicon (MMSL)*; *NANDA nursing diagnoses*; *NCBI Taxonomy* a mnoho dalších.

3. Konverzní nástroje

Rostoucí počet klasifikačních systémů a nomenklatur si vyžádal vytváření konverzních nástrojů mezi hlavními systémy při přenosu informací mezi různými datovými bázemi. Modelují se rozsáhlé ontologie a sémantické sítě pro přenos mezi různými databázemi. Jsou navrhovány různé metatezaury, které monitorují a spojují informace z různých heterogenních zdrojů.

Vývoj UMLS (**U**nified **M**edical **L**anguage **S**ystem) [21] začal v roce 1986 v Národní lékařské knihovně v USA. UMLS znalostní zdroje jsou univerzální, to znamená, že nejsou optimalizované pro jednotlivé aplikace. Jedná se o inteligentní automatizovaný systém, který „rozumí“ biomedicínským termínům a jejich vztahům a využívá tohoto porozumění ke čtení a organizování informací ze strojově zpracovatelných zdrojů. Jeho cílem je kompenzace terminologických a kódových rozdílů těchto nesourodých systémů a současně i jazykových variací uživatelů. Jedná se o vícejazyčný slovník klasifikačních systémů na velkokapacitním médiu, což umožňuje převod kódovaných termínů mezi různými klasifikačními systémy.

UMLS se skládá ze tří znalostních zdrojů: Metathesaurus, Semantic Network a SPECIALIST lexicon. Semantic Network obsahuje informace o sémantických druzích a jejich vztazích. Ve SPECIALIST lexikonu každé slovo nebo termín zaznamenává syntaktické, morfologické a ortografické informace.

UMLS Metathesaurus je rozsáhlá, víceúčelová a vícejazyčná lexikonová databáze, která zahrnuje informace o biomedicínských, zdravotnických a jim příbuzným pojmech, obsahuje jejich různé názvy a vztahy mezi nimi. Jeho hlavním cílem je spojení alternativních názvů stejných pojmů a identifikování užitečných vztahů mezi různými pojmy. Pokud různé slovníky používají stejný název pro různé termíny, v *Metatezauru* se objeví oba významy a ukáže se v něm, který význam je použit v kterém slovníku. Pokud se stejný termín objevuje v různých slovnících v různých hierarchických kontextech, v *Metatezauru* jsou zachyceny všechny tyto hierarchie. *Metatezaurus* nepodává jeden konzistentní pohled, ale zachovává mnoho pohledů, které jsou obsaženy ve zdrojových slovnících.

4. Minimální datový model pro kardiologii

Od roku 2000 se naše oddělení věnovalo výzkumu na společném pracovišti dvou univerzit, dvou nemocnic a Ústavu informatiky AV ČR. Toto společné pracoviště nese název EuroMISE centrum a spravuje také dvě ambulance preventivní kardiologie.

V roce 2002 jsme vyvinuli Minimální datový model pro kardiologii (MDMK) [22], [23]. MDMK je soubor přibližně 150 atributů, jejich vzájemných vztahů, integritních omezení, jednotek, atd. Přední odborníci na poli české kardiologie se shodli na těchto attributech jako na základních datech, která jsou nezbytná pro vyšetření pacienta na kardiologii.

Jelikož je kardiologie velice rozsáhlý obor, byl MDMK zaměřen pouze na aterosklerotická kardiovaskulární onemocnění. Cílem tohoto datového modelu bylo vytvoření minimálního souboru znaků, které je potřeba sledovat u pacientů z hlediska aterosklerotického kardiovaskulárního onemocnění, aby mohl být pacient následně zařazen mezi osoby nemocné či rizikové.

MDMK se skládá z několika skupin znaků. Nezbytné administrativní atributy tvoří jednu část MDMK. Nicméně, ty nejsou zahrnuty do naší analýzy, jelikož analyzované lékařské zprávy byly anony-

mizované a tudíž bez administrativních dat.

MDMK se skládá z osmi skupin atributů. Po *administrativní* části následuje *rodinná anamnéza* zahrnující informace o matce, otci a libovolném počtu sourozenců. Dále následuje *sociální anamnéza a toxikomanie*, která se zaměřuje na rodinný stav, fyzickou zátěž, psychickou zátěž, fyzické aktivity, míru kouření a míru požívání alkoholu. Část MDMK je věnována *alergiím* pacienta, zejména alergiím na léky. V části *osobní anamnézy* je zjišťována přítomnost diabetu mellitu, hypertenze, hyperlipoproteinémie, ischemické choroby srdeční a její konkrétní formy, je zjišťováno, zda pacient prodělal cévní mozkovou příhodu, zda se léčí s ischemickou chorobou periferních tepen, jsou zde atributy týkající se aneurysma aorty, ostatních relevantních chorob a u žen menopauzy. V části MDMK nazvané *Současné obtíže možného kardiálního původu* se lékaři zaměřují na dušnost, bolest na hrudi, palpitace, otoky, synkopu, kašel, hemoptýzu a klaudikaci. Další část MDMK zjišťuje, jakou *léčbu* pacient podstupuje, jaký má předepsaný druh diety a jaké užívá léky. V části *fyzikálních vyšetření* se zjišťuje pacientova hmotnost, výška, tělesná teplota, obvod boků, BMI, WHR, krevní tlak, tepová a dechová frekvence a patologické nálezy. *Laboratorní vyšetření* se zaměřují na glykémii, kyselinu močovou, celkový cholesterol, HDL-cholesterol, LDL-cholesterol a triacylglyceroly. Poslední část MDMK tvoří atributy vztahující se k EKG, kde se zjišťuje rytmus, frekvence, průměrné intervaly PQ a QRS a je zde prostor pro celkový popis EKG.

Na základě MDMK byla vytvořena softwarová aplikace ADAMEK (Aplikace Datového Modelu EuroMISE centra Kardio), kde jsou data ukládána strukturovaně. Po jejím dokončení byl od března 2002 zahájen sběr dat. Data byla a stále jsou sbírána ve dvou ambulancích preventivní kardiologie EuroMISE centra. K 1. dubna 2010 zde byla data o 1189 pacientech.

5. Jazyk česky psaných lékařských zpráv a aplikace mezinárodních klasifikačních systémů v MDMK

5.1 Český jazyk

Český jazyk patří k západní skupině slovanských jazyků. Mezi slovanskými jazyky patří do východní, nebo-li satemové, skupiny indoevropských jazyků. Český jazyk se od ostatních slovanských jazyků odlišil několika změnami, které proběhly v rozmezí 10. až 16. století. Ke konci 15. století ztratila čeština prakticky duální číslo a dva ze slovanských minulých časů. Na druhou stranu ale významněji rostl mluvený aspekt a zvýšil se počet skloňování. Na začátku 15. století navrhnul církevní reformátor Jan Hus diakritický pravopis. Umístil diakritická znaménka nad některá písmena latinky a tím odlišil české palatální souhlásky (č, ě, ě, ř, š, ť, ž) a dlouhé samohlásky (á, é, í, ó, ú, ý). V 16. století bylo ještě přidáno písmeno „ů“, které označovalo dlouhé „u“. Jediné zdvojené písmeno v moderní češtině je „ch“.

Česky mluví zhruba 10 milionů lidí v České republice a asi 200 000 v dalších zemích. Jedná se většinou o emigranty a jejich děti, kteří opustili zemi ve značných migracích kolem 1. a 2. světové války a kolem roku 1948 a 1968. Mnoho česky mluvících obyvatel žije zejména v Rakousku (většinou ve Vídni), v Polsku, Německu, na Ukrajině, v Chorvatsku, v západním Rumunsku, v Austrálii a v Kanadě. Nicméně, největší skupina česky mluvících lidí mimo Českou republiku žije ve Spojených státech amerických, ve městech jako New York, Chicago a v několika venkovských komunitách v Texasu, Wisconsinu, Minnesotě a Nebrasce.

Čeština patří mezi jazyky s volným slovosledem [24].

5.2 Jazyková analýza česky psaných lékařských zpráv ve volném textu

Styl zapisování textových zpráv není nijak standardizován. Rozdíly najdeme nejenom ve zprávách od různých lékařů, ale i jednotliví lékaři často zapisují stejné koncepty v různých tvarech. Následující část je zaměřena na již zmíněné jazykové

a lexikální rozdíly v českých lékařských zprávách psaných volným textem.

Diakritika: Jak jsme se již dříve zmínili, český jazyk využívá diakritický pravopis. Jako příklad diakritických písmen uveďme například písmena „ě, č, ř, ž“ a další. Nicméně, pro lékaře je rychlejší psát bez diakritických znamének a použít písmena „e, c, r, z“. Takový text je potom pro rodilého Čecha srozumitelný, ale je těžký pro počítačové zpracování.

Překlepy: Větším problémem jsou překlepy, které jsou velmi časté a text je potom dále velmi těžce použitelný pro počítačové zpracování.

Mezery: Podobnou záležitostí je i vynechávání mezer mezi slovy, kdy se ze dvou slov stává jedno slovo a to je opět nepoužitelné pro počítačové zpracování. Lékaři se různí v zapisování mezer před jednotkami. Můžeme se setkat jak s tvarem s mezerou, např. „2,5 mg“, tak i s tvarem bez mezery, např. „4mg“. Tak to je i s tvary, kde se používá lomítko. Někteří lékaři používají variantu bez mezer, např. „80/min“, jiní variantu s mezerami „70/min“.

Číslice 0: Pro počítačové zpracování je také složité, když někteří lékaři používají místo číslice 0 velké písmeno O.

Zkratky: Jelikož lékaři mívají málo času na zapisování zpráv, dochází ke zkracování slov. Problémem je, že neexistuje žádné pravidlo, jak by daný atribut měl být zkrácen. Proto se můžeme setkat s různě zkrácenými stejnými slovy. Může se stát, že v jedné a té samé zprávě je slovo zkráceno dvakrát a pokaždé jinak. Se zkrácenými tvary souvisí také to, že se setkáváme s vynecháním tečky za zkráceným slovem.

Zaokrouhlování: Další část, ve které můžeme nalézt mnoho rozdílů, souvisí s číselnými hodnotami. Zde se můžeme například setkat u stejného znaku u jednoho lékaře se zaokrouhlováním hodnot na celá čísla, u jiného lékaře s uváděním hodnot nezaokrouhlených, s přesností na jedno nebo dvě desetinná čísla. Někdy jsou číselné hodnoty znaku uváděné jako rozmezí, např. „70-80“. Častokrát bývá zadán pouze přibližný

údaj, například „diastolický tlak kolem 70“. U některých znaků nejsou hodnoty vyjádřeny číslem, ale pouze slovně, např. „tlak je zcela v mezích normy“.

Římské a arabské číslice: Rozdíl je i v používání římských a arabských číslic. Například u zápisu o srdečních ozvách lze najít jak tvar „ozvy 2“, tak i „ozvy II“.

Synonyma: Český jazyk je velmi bohatý na synonyma a ta nacházíme i v lékařských zprávách.

Pravopis: Někteří lékaři používají starší formy pravopisu, někteří novější.

Časové údaje: Ani zaznamenávání časových údajů není sjednoceno. V lékařských zprávách se objevuje jak název měsíce, např. „únor 2006“, tak i pořadí měsíce, např. „2/2006“.

Podávání léků: V lékařských zprávách se objevuje hodně způsobů, jak popsat, kdy by si měl pacient vzít lék (např. 1-0-0 vs. 1 tabl. ráno vs. 1 ráno vs. 1x ráno).

Hodnoty znaků: Často jsou stejné hodnoty znaku zapisovány řadou různých způsobů. Například hodnota znaku diabetes mellitus bývá zapsána jako: diabet, diabet., diabetes mellitus 2. typu, diabetička 2. typu na dietě, diabetes mellitus II. typu na dietě, DM 2. typu, DM 2. typu.

Nejedná se ale jenom o problém při zapisování lékařských zpráv, ale stejné chyby můžeme najít např. i na webových stránkách [25].

5.3 Analýza znaků Minimálního datového modelu pro kardiologii v textových lékařských zprávách

V analýze textových lékařských zpráv jsem vycházela ze znaků Minimálního datového modelu pro kardiologii. Lékařské zprávy byly anonymizované a z tohoto důvodu nebylo možné analyzovat administrativní data.

Když se podíváme na jednotlivé znaky, tak pouze *diastolický* a *systolický tlak* jsou zaznamenány ve všech textových lékařských zprávách, které byly analyzovány. V 96,30 % textových lékařských zprávách jsou zaznamenány *léky*, které pacient užívá nebo které lékař nově

předepisuje. I hmotnost je zaznamenána v 96,30 % zpráv. Oproti tomu *výška* už pouze v 74,07 % zpráv. V analyzovaných textových zprávách se některé znaky MDMK neobjevily ani v jednom případě. Jedná se například o tyto znaky: *aneurysma aorty*, *angína pectoris*, *ICHDK*, *ischemická CMP*, *kdy zjištěn DM*, *kdy zjištěna HLP*, *manifestní onemocnění periferních tepen (jiných než ICHDK)*, *medikamentózní léčba HPL*, *němá ischemie*, *sinuskomorová blokáda a tělesná teplota*. *Alergie na lék* je zmíněná ve 22,22 % zpráv, to, zda pacient pije nebo nepije *alkohol* v 51,85 % zpráv, bolesti na hrudi v 37,04 %. *Celkovou psychickou zátěž* zaznamenává 11,11 % zpráv, *fyzickou zátěž v zaměstnání* 11,11 % zpráv, *celkový cholesterol* 70,37 %, *pítí černé kávy* 22,22 %, *různá další vyšetřování* 62,96 %. *Dechová frekvence* byla nalezena ve 3,7 % zpráv, *diabetes mellitus* ve 40,74 %, *dieta* u 59,26 %, *glykémie* u 51,85 %, *HDL cholesterol* u 66,67 %. Se zaznamenáním přítomnosti/nepřítomnosti *hypertenze* jsme se mohli setkat u 70,37 %, *hypertrofie levé komory* u 11,11 %, *s infarktem myokardu* u 14,81 %, *s průměrným množstvím cigaret u kuřáka* u 51,85 % a tak dále.

Lze tedy říci, že při zapisování výsledků vyšetření pomocí volného textu zůstává plno znaků nezaznamenáno. K tomu může docházet z několika důvodů. Lékaři nemají přesně danou osnovu, podle které by měli postupovat a může se stát, že na některé znaky mohou zapomenout. V softwarových aplikacích je tomuto problému zabráněno tak, že pokud lékař nevyplní žádanou položku, program mu nedovolí pokračovat. Dalším důvodem, proč nejsou některé znaky v textové zprávě zaznamenány, může být fakt, že lékařům ze znalosti předchozích znaků vyplyne, že další znak nemůže být přítomen a proto se již na něj dále nezeptají a nezaznamenají ho. Z textové zprávy ale nevyplyne, zda skutečně byly u pacienta zjišťovány tyto základní informace, z jejichž hodnot lékaři hodnoty dalších znaků sami svými znalostmi vyvodili.

5.4 Analýza znaků Minimálního datového modelu pro kardiologii v softwarové aplikaci ADAMEK

Pro analýzu bylo využito 1118 lékařských zpráv z ambulance preventivní kardiologie EuroMISE centra.

Ve všech lékařských zprávách, tedy ve 100 %, byla zaznamenána například *alergie na lék*, *aneurysma aorty*, *angína pectoris*, *bolest na hrudi*, *otoky dolních končetin*, *dušnost*, *hypertrofie levé komory*, *infarkt myokardu*, *ischemická CMP*, *jiná alergie*, *kašel po ACE inhibitorech*, *klaudikace*, *němá ischemie*, *systolický tlak*, *typ léčby diabetu*. Již z tohoto výčtu je patrné, že pomocí softwarové aplikace je zaznamenáváno daleko větší množství znaků.

Při porovnání zápisu jednotlivých znaků v textové a strukturované lékařské zprávě, dojdeme například k těmto výsledkům. *Alergie na lék* je v textových lékařských zprávách zaznamenána ve 22,2 %, v aplikaci ADAMEK ve všech, tedy 100 %, zpráv. Odpověď, zda pacient má nebo nemá *aneurysma aorty*, byla v aplikaci ADAMEK vyplněna ve všech zprávách, ale v žádné textové lékařské zprávě. Na *bolest na hrudi* se lékaři, vyplňující lékařskou zprávu pomocí aplikace ADAMEK, zeptali ve všech případech, v textových lékařských zprávách pouze 37,0 % obsahuje zmínku o bolesti na hrudi. *Celková psychická zátěž* byla vyplněna u 96,2 % zpráv v aplikaci ADAMEK, v textových lékařských zprávách to bylo 11,1 %. Stejného procenta zaznamenání v textových lékařských zprávách bylo dosaženo u znaku *fyzická zátěž v zaměstnání*, v aplikaci ADAMEK to bylo v 94,8 %. *Celkový cholesterol* byl ve zprávách aplikace ADAMEK uveden u 83,4 % zpráv, v textových lékařských zprávách to bylo v 70,4 %. Přítomnost nebo nepřítomnost *diabetu mellitu* byla v softwarové aplikaci vyplněna u 95,9 % lékařských zpráv, u textových lékařských zpráv to bylo 40,7 %. *Glykémie* je v aplikaci ADAMEK uvedena u 77,6 % zpráv, v 51,9 % v textových lékařských zprávách. *Cholesterol* je zaznamenán v softwarové aplikaci v 917 zprávách, což je v 82,0 %, zatímco v textových lékařských zprávách se se zaznamenáním tohoto znaku můžeme setkat v 66,7 % analyzovaných zpráv. Nejblíže k sobě mají oba druhy

zpráv u hmotnosti, která je v aplikaci ADAMEK zapsána u 97,9 % zpráv a v textových lékařských zprávách v 96,3 %. Se zaznamenáním přítomnosti či nepřítomnosti *hypertenze* se setkáme v 95,3 % zpráv z aplikace ADAMEK, zatímco v textových lékařských zprávách pouze v 70,4 % případů. Velký rozdíl nalezneme například u *hypertrofie levé komory*, která je v lékařských zprávách aplikace ADAMEK zapsána ve všech případech, ale pouze v 11,1 % textových lékařských zpráv nebo *menopauza*, která je v aplikaci ADAMEK uvedena v 96,9 % zpráv a pouze v 7,4 % v textových lékařských zprávách. *Interval PQ* je v lékařských zprávách zapsaných pomocí softwarové aplikace zaznamenán v 89,1 % a v textových lékařských zprávách v 62,9 %, obdobně i *interval QRS* je ve strukturovaných zprávách aplikace ADAMEK zaznamenán v 89,5 % a v textových lékařských zprávách v 66,7 %.

V Tabulce 1 je přehledně zobrazeno procentuální vyjádření o zaznamenaných hodnotách vybraných znaků MDMK v lékařských zprávách při užití softwarové aplikace ADAMEK a v textových lékařských zprávách.

5.5 Atributy Minimálního datového modelu pro kardiologii zakódované pomocí SNOMED CT

Tabulka 2 ukazuje několik příkladů atributů z Minimálního datového modelu, kterým bylo přiděleno ConceptID z klasifikačního systému SNOMED CT. Prvním předpokladem kódování, je ale přeložení názvu atributů do anglického jazyka, jelikož v současné době stále neexistuje česká verze.

5.6 Atributy Minimálního datového modelu pro kardiologii zakódované pomocí MKN-10

Jelikož je Mezinárodní klasifikace nemocí jednou z mála mezinárodních medicínských klasifikací, které jsou přeložené do českého jazyka, pokusila jsem se zakódovat termíny Minimálního datového modelu právě pomocí této klasifikace. V Tabulce 3 jsou pro srovnání uvedeny rovněž kódy atributů MDMK v systému SNOMED CT.

Tab. 1. Procentuální vyjádření zaznamenaných hodnot vybraných znaků MDMK v 1118 lékařských zprávách při užití aplikace ADAMEK

znak MDMK	aplikace ADAMEK n=1118	textové lékařské zprávy n=27	textové lékařské zprávy – 95% interval spolehlivosti	
			dolní mez	horní mez
alergie na lék	100,0 %	22,2 %	8,6 %	42,3 %
aneurysma aorty	100,0 %	0,0 %	0,0 %	12,8 %
angína pectoris	100,0 %	0,0 %	0,0 %	12,8 %
bolest na hrudi	100,0 %	37,0 %	19,4 %	57,6 %
celková psychická zátěž	96,2 %	11,1 %	2,4 %	29,2 %
celkový cholesterol	83,4 %	70,4 %	49,8 %	86,2 %
diabetes mellitus	95,9 %	40,7 %	22,4 %	61,2 %
dušnost	100,0 %	55,6 %	35,3 %	74,5 %
fyzická zátěž v zaměstnání	94,8 %	11,1 %	2,4 %	29,2 %
glykémie	77,6 %	51,9 %	32,0 %	71,3 %
HDL cholesterol	82,0 %	66,7 %	46,0 %	83,5 %
hmotnost	97,9 %	96,3 %	81,0 %	99,9 %
hypertenze	95,3 %	70,4 %	49,8 %	86,2 %
infarkt myokardu	100,0 %	14,8 %	4,2 %	33,7 %
interval PQ	89,1 %	62,9 %	42,4 %	80,6 %
jiná alergie	100,0 %	18,5 %	6,3 %	38,1 %
klaudikace	100,0 %	3,7 %	0,9 %	19,0 %
kuřák	96,5 %	66,7 %	46,0 %	83,5 %
tělesná teplota	14,9 %	0,0 %	0,0 %	12,8 %

Tab.2. Vybrané atributy MDMK zakódované pomocí SNOMED CT.

Atributy z MDMK	English equivalent	SNOMED CT (Concept ID)
alergie na léky	Drug allergy (disorder)	416098002
	Allergic reaction to drug (disorder)	416093006
hypertenze	Essential hypertension (disorder)	59621000
	High blood pressure (&essential hypertension)	194757006
	Essential hypertension NOS (disorder)	266228004
ischemická choroba srdeční	Ischemic heart disease (disorder)	414545008
dušnost	Asthma (disorder)	187687003
bolest na hrudi	Dull chest pain (finding)	3368006
palpitace	(Palpitations) or (awareness of heartbeat) or (fluttering of heart)	161965005
otoky	Swelling or edema (finding)	248477007
synkopa	Syncope (disorder)	271594007
klaudikace	Claudication (finding)	275520000
hmotnost	On examination – weight NOS (finding)	162770007
	Height and weight (observable entity)	162879003
výška	Body height measure (observable entity)	50373000
tělesná teplota	Body temperature finding	105723007
	Body temperature (observable entity)	276535009
obvod pasu	Abdominal girth measurement	48094003

Tab. 3. Vybrané atributy MDMK zakódované pomocí MKN-10 a SNOMED CT.
1. část – Alergie

Atributy z MDMK	Termín v MKN 10	Kód MKN 10	English equivalent	SNOMED CT (Concept ID)
alergie přítomna	alergie	T78.4	allergy manifested	nenalezeno
alergie na léky	alergie na lék	T88.7	drug allergy (disorder)	416098002
			allergic reaction to drug (disorder)	416093006

Tab. 3. Vybrané atributy MDMK zakódované pomocí MKN-10 a SNOMED CT.
2. část – Osobní anamnéza

Atributy z MDMK	Termín v MKN 10	Kód MKN 10	English equivalent	SNOMED CT (Concept ID)
diabetes mellitus	diabetes typu I	E10.-	diabetes mellitus type 1 (disorder)	46635009
	inzulin dependentní	E10.-	insulin-treated non-insulin- dependent diabetes mellitus (disorder)	237599002
hypertenze	Esenciální (primární) hypertenze	I10	essential hypertension (disorder)	59621000
ischemická choroba srdeční	ischemie koronární	I25.9	ischemic heart disease (disorder)	414545008

Jak ze samotného názvu Mezinárodní klasifikace nemocí vyplývá, je možné tuto klasifikaci použít zejména pro zakódování nemocí, syndromů, patologických stavů, poranění, obtíží a jiných důvodů pro styk se zdravotnickými službami, tj. toho typu informací, které bývají registrovány lékařem. Bohužel, pomocí této klasifikace tedy nemůžeme zakódovat řadu atributů Minimálního datového modelu pro kardiologii, jako např. rodinný stav, vzdělání, psychickou zátěž, fyzickou zátěž, tělesnou aktivitu, kouření, pití alkoholu, fyzikální vyšetření (hmotnost, výška, tělesná teplota, obvod pasů, obvod boků, BMI, WHR, atd.), laboratorní vyšetření (celkový cholesterol, HDL-cholesterol) a ani popis EKG. MNK se hodí pouze pro části Minimálního datového modelu pro kardiologii týkající se osobní anamnézy a pro současné potíže možného kardio-vasculárního původu (viz Tabulka 3).

5.7 Standardizace klinického obsahu

Atributy z pohledu možnosti jejich mapování na standardní kódovací systémy lze klasifikovat následujícím způsobem:

- *Bezproblémové atributy*, tj. atributy, které lze mapovat přímým způsobem, tak, že u nich existuje právě jedna možnost mapování, případně existují pouze synonyma se zcela stejným významem a tedy i klasifikačním kódem (např.: křestní jméno pacienta, současný kuřák, hybnost, výška pacienta).
- *Částečně problematické atributy*, tj. atributy, které lze mapovat tak, že u nich existuje právě několik možností mapování na různá synonyma, která se ale lehce liší významem a tedy zpravidla i klasifikačním kódem (např.: ischemická cévní mozková příhoda, angína pectoris, hypertenze, městnavé srdeční selhání).
- *Atributy s příliš malou granularitou*, tj. atributy popisující určitou vlastnost na příliš obecné úrovni tak, že klasifikační systémy obsahují pouze termíny užšího významu (např. email v MDMKP vs. email do zaměstnání / email domů / email lékaře atd. v klasifikačních systémech).

- *Atributy s příliš velkou granularitou*, tj. atributy popisující určitou vlastnost na úzké úrovni tak, že klasifikační systémy obsahují pouze termíny obecnějšího významu (např. souměrný tep karotid, atd.).
- *Atributy, které se v klasifikačních systémech dohledat nedaří*, např. dyslipidemie, atd.

Pro řešení takovýchto mapovacích problémů je důležitá úzká spolupráce s lékaři. Často je potřeba vybrat správné synonymum, které by nahradilo určitý odborný termín. Je ale důležité provést to opatrně, aby se informace neztratila nebo aby nedošlo ke špatné interpretaci. V případě, že to nejde udělat bez ztráty informace, je lepším způsobem popsat termín bez kódu pomocí několika kódovaných termínů, eventuálně s ukázkou jejich vzájemných sémantických vztahů. Jestliže toto není možné, můžeme polemizovat se specialisty, jestli tyto „nepopsatelné“ termíny (atributy) mohou být nahrazeny jiným, ekvivalentnějším nebo standardnějším termínem. Ve speciálních případech je možné přidat určitý termín do připravované nové verze určitého kódovacího systému. V případě, že není možné použít žádnou z výše zmíněných možností, které by řešily mapovací problémy, je nezbytné se vyrovnat s tím, že mapování nebude nikdy 100%. Nedostatečný mapující proces limituje interoperabilitu heterogenních systémů, které se ve zdravotnictví používají pro různé účely. Omezená interoperabilita je často nevyhnutelná z vlastního kořene problému, tj. nedostatečná harmonizace klinických kontextů heterogenních systémů elektronických zdravotních záznamů.

6. Závěr

Analýzou textových lékařských zpráv bylo zjištěno, že zapisování pomocí volného textu je velice nehomogenní a nestandardizované. Největšími problémy pro další počítačové zpracování jsou překlady, různá délka zkracovaných výrazů a používání synonym. Standardizovaná terminologie by přinesla výhody lékařům, pacientům, administrátorům, softwarovým vývojářům a plátcům. Standardizovaná klinická terminologie by pomohla poskytovatelům lékařské péče tak, že by jim poskytla jednodušeji dostupné a kompletní

informace, které náleží k procesu zdravotnické péče (chorobopis pacienta, nemoci, léčby, laboratorní výsledky, atd.) a to by vyústělo v lepší výsledky v péči o pacienta.

I přes problémy, které při využití mezinárodních nomenklatur a metate-zaurů ve zdravotnictví v České republice přetrvávají, je jejich využití prvním nezbytným krokem k umožnění interoperability heterogenních systémů zdravotních záznamů. Dostatečná sémantická interoperabilita těchto systémů je základem pro sdílenou zdravotní péči, která vede k efektivitě ve zdravotnictví, finančním úsporám i snížení zátěže pacientů, a proto se ve své práci snažím analyzovat, jak mezinárodních klasifikačních systémů využít co nejlépe pro potřeby českého zdravotnictví.

Současné zdravotnické informační systémy umožňují sbírat různé klinické informace, tyto systémy jsou propojeny s klinickými znalostními databázemi, mohou vyhledávat data, shromažďovat data, analyzovat data, vyměňovat si data a mají i plno dalších funkcí. Jako nejlepším klasifikačním systémem se zatím jeví SNOMED CT, který může poskytnout základy pro tyto funkce. Informační systémy mohou využít koncepty, hierarchie a vztahy jako společný referenční bod. SNOMED CT ale může i překročit přímou péči o pacienty. Tato terminologie může, například, usnadnit podporu rozhodování, statistické zpracovávání, sledování veřejného zdraví, zdravotnický výzkum a analýzy nákladů.

Poděkování

Práce byla podpořena projekty SVV-2010-265513; 1M06014 MŠMT ČR a AV0Z 10300504 ÚIAV ČR, v.v.i.

Literatura

- [1] Bleich H. L., Slack W. V.: Reflections on electronic medical record: When doctor will use them and when they will not, *Int. J. Med. Inform.* 79(2010) 1-4.
- [2] Hoerbst A., Kohl C. D., Knaup P., Ammenwerth E.: Attitudes and behaviors related to the introduction of electronic health records among Austrian and German citizens, *Int. J. Med. Inform.* 79(2010) 81-89.
- [3] Rinner C., Janzek-Hawlat S., Sibinovic S., Duftschmid G.: Semantic Validation of Standard-based Electronic Health Record

Documents with W3C XML Schema.

Method Inf Med 49 (2010), preprint online

- [4] Oemig F., Blobel B.: Semantic Interoperability Adheres to Proper Models and Code Systems: A Detailed Examination of Different Approaches for Score Systems. *Methods Inf Med* 2010 49 2, 148-155
- [5] Lopez D.M., Blobel B.: A development framework for semantic interoperable health information systems. *Int J Med Inform* 2009; 78 (2), 83-103
- [6] Garde S., Knaup P., Hovenga E.J.S., Heard S.: Towards Semantic Interoperability for Electronic Health Records: Domain Knowledge Governance for openEHR Archetypes. *Methods Inf Med* 2007; 46 (3), 332-343
- [7] Nagy M., Hanzlíček P., Přečková P., Kolesa P., Mišúr J., Dioszegi M., Zvárová J.: Building Semantically Interoperable EHR Systems Using International Nomenclatures and Enterprise Programming Technique. In *eHealth: Combining Health Telematics, Telemedicine, Biomedical Engineering and Bioinformatics to the Edge*. Amsterdam: IOS Press, 2008 (Eds. Blobel, B.; Pharow, P.; Zvárová, J.; Lopez, D.) 105-110
- [8] Nagy M., Hanzlíček P., Přečková P., Říha A., Dioszegi M., Seidl L., Zvárová J.: Semantic Interoperability in Czech Healthcare Environment Supported by HL7 version 3. *Methods Inf Med* 2010 49 (2), 186-195
- [9] Zvárová J., Hanzlíček P., Nagy M., Přečková P., Zvára K., Seidl L., Bureš V., Šubrt D., Dostálová T., Seydlová M.: Biomedical Informatics Research for Individualized Life-long Shared Healthcare. In: *Biocybernetics and Biomedical Engineering*, 2009, 29 (2), 31-41
- [10] Přečková P., Špidlen J., Zvárová J.: Usage of the International Nomenclatures and Metathesauruses in Shared Healthcare in the Czech Republic. *Acta Informatica Medica*, 2005 (13), 201-205
- [11] Přečková P., Zvárová J., Špidlen J.: International Nomenclatures in Shared Healthcare in the Czech Republic. In: *Proceedings of 6th Nordic Conference on eHealth and Telemedicine „From Tools to Services“* (Ed.: Doupi P.), 2006, 45-46
- [12] Elkin P. L., Trusko B. E., Koppel R., Speroff T., Mohrer D., Sakji S., Gurewitz I., Tuttle M., Brown S. H.: Secondary Use of Clinical Data. In *Seamless Care Safe Care*. IOS Press, 2010 (Eds. Blobel B., Hvannberg E., Gunnarsdóttir), 14-29
- [13] Stausberg J., Lehmann N., Kaczmarek D., Stein M.: Reability of diagnose coding with ICD-10. *Int. J. Med. Inform.* 2008 (77), 50-57
- [14] IHTSDO: The International Health Terminology Standards Development Organization: SNOMED Clinical Terms® User Guide 2008

- [15] Cornet R.: Definitions and Qualifiers in SNOMED CT. *Methods Inf Med* 2009 (48), 177-183
- [16] Schulz S., Hanser S., Hahn U., Rodgers J.: The Semantics Procedures and Diseases in SNOMED® CT. *Methods Inf Med* 2006 (45), 354-8
- [17] Gault Lora V., Schultz M.: Variations in Medical Subject Headings (MeSH) mapping: from the natural language of patron terms to the controlled vocabulary of mapped lists. *J Med Libr Assoc.* 2002 April; 90(2): 173180
- [18] Khan A. N., Griffith S. P., Moore C., Russell D., Rosario A. C., Jr., Bertolli J.: Standardizing Laboratory Data by Mapping to LOINC. *J Am Med Inform Assoc.* 2006 MayJun; 13(3): 353355
- [19] Louis D. N., Ohgaki H., Wiestler O. D., Cavenee W. K., Burger P.C., Jouvett A., Scheithauer B.W., Kleihues P.: The 2007 WHO Classification of Tumours of the Central Nervous System. *Acta Neuropathol.* 2007 August; 114(2): 97109
- [20] Brierley J.: The evolving TNM cancer staging system: an essential component of cancer care. *CMAJ.* 2006 January 17; 174(2): 155156
- [21] Campbell J.R., Olivek D.E., Shortliffe: UMLS: towards a collaborative approach for solving terminological problems, *J. Am. Med. Inform. Assoc.* 5 (1998), 12-16
- [22] Adášková J., Anger Z., Asche-rmann M., Bencko V., Berka P., Filipovský J., Goláš L., Grus T., Grünfeldová H., Haas T., Hanuš P., Hanzlíček P., Holcátová I., Hrach K., Jiroušek R., Kejřová E., Kocmanová D., Kolář J., Kotásek P., Králíková E., Krupařová M., Kylvoušková M., Malý M., Mareš R., Matoulek M., Mazura I., Mrázek V., Novotný L., Novotný Z., Pecen L., Peleška J., Prázný M., Pudil P., Rameš J., Rauch J., Reissigová J., Rosolová H., Rousková B., Říha A., Sedlak P., Slámová A., Somol P., Svačina Š., Svátek V., Šabík D., Šimek S., Škvor J., Špidlen J., Štochl J., Tomečková M., Umnerová V., Zvára K., Zvárová J.: A proposal of the Minimal Data Model for Cardiology and the ADAMEK software application (in Czech). Internal research report of the EuroMISE Centre Cardio. Institute of Computer Science AS CR. Prague, October 2002.
- [23] Mareš R., Tomečková M., Peleška J., Hanzlíček P., Zvárová J.: Interface of patient database systems - an example of the application designed for data collection in the framework of Minimal Data Model for Cardiology (in Czech). *Cor et Vasa*, 2002, 44 (4), Suppl., 76
- [24] Eryigit G., Nivre J., Oflazer K.: Dependency Parsing of Turkish. *Computational Linguistics.* 2008, 34(3), 357-389.
- [25] Ringlestetter C., Schulz K. U., Mihov S.: Orthographic Errors in Web Pages: Toward Cleaner Web Corpora. *Computational Linguistics.* 2006, 32(3) 295-340

Kontakt

Mgr. Petra Přechová

Centrum biomedicínské informatiky

Ústav informatiky AV ČR, v.v.i.

Pod Vodárenskou věží 2

182 07 Praha 8

e-mail: preckova@euromise.cz

Porovnávání internetových stránek a lékařských doporučených postupů s využitím databází řízených medicínských slovníků v kontextu hodnocení kvality webových zdrojů

D. Rak¹

¹1. lékařská fakulta UK v Praze

Školitel: Doc. Ing. Vojtěch Svátek, Dr.

Souhrn

IModerní technologie poskytují řadu možností jak publikovat na internetu téměř cokoli, co autor zamýšlí. To se týká i medicínských informací, kde se široká dostupnost mnoha informací s takto závažným obsahem ve velmi variabilní kvalitě, stává velkým problémem. Bohužel, v současné době jediná reálná možnost, jak ověřit kvalitu medicínských textů, je verifikace odborníkem v oboru. Doporučené postupy (DP) v lékařství jsou systematicky vytvářeny a publikovány renomovanými lékařskými společnostmi na základě výsledků medicíny založené na důkazech (MZD).

Cílem této práce je předložit jednoduchý postup posouzení webových stránek (WS) s využitím řízených medicínských slovníků a DP, jakožto standardu pro hodnocení kvality obsahu. Postup je založen na myšlence, že sdělení dokumentu nebo přinejmenším jeho informační rozsah je reflektován použitými odbornými výrazy v textu. Diskutujeme možné využití tohoto přístupu v procesu poloautomatického hodnocení kvality včetně dalších souvislostí jeho aplikace.

Podstatou přístupu je vyhledání kandidátů na koncepty v DP a v testovaných WS prostřednictvím mapování na UMLS. Tím jsou získány seznamy obsažených termínů a jim odpovídajících konceptů. Pro DP byly navrženy a vyhodnoceny nejvhodnější agregace, proti nimž byla následně analyzována podobnost testovaných WS na úrovni termínů a konceptů. Metoda byla aplikována na vybranou medicínskou oblast s využitím tří relevantních DP a 100 WS. Analyzované WS spadaly do pěti odlišitelných kategorií (s ohledem na jejich cílovou skupinu). Byly navrženy a vyhodnoceny agregace termínů pro DP. Průměrná kosinová podobnost všech WS k DP dosáhla 0,69, přičemž

podobnostní průměry za kategorie WS variovaly v rozmezí 7,6 % oproti celkovému průměru.

Dosavadní výsledky jsou prvním krokem směrem k automatickému hodnocení medicínských webových zdrojů s využitím DP jakožto standardu kvality. Diskutujeme další plánované úlohy, mířené k vylepšení výstupů porovnání a jejich obecnosti.

Klíčová slova: hodnocení informační kvality, řízené medicínské slovníky, UMLS, medicína založená na důkazech (MZD), lékařské doporučené postupy (DP), informační kvalita, anotace, podobnost, koncept, reprezentace obsahu

1. Úvod

S mohutným rozvojem informačních technologií se stává velmi snadným publikování v podstatě jakýchkoli informací kýmkoli veřejně na internetu. V dnešní době jsou široké veřejnosti k dispozici jednak již klasické metody vytváření statických i dynamických webových stránek a jejich následné publikování. I když nedostatečně, v této podobě je publikovaná informace alespoň částečně svázaná s tím, kdo obsah vytváří, případně s tím, na koho je příslušná doména registrována. Nově se také objevila řada dalších možností, označovaných pojmem „Web 2.0“, kdy je do tvorby veřejně dostupného obsahu zapojen přímo uživatel či čtenář stránky. Jde zejména o různé blogy, wiki systémy, sociální sítě, diskusní skupiny a podobně. Za obsah pak v podstatě neručí nikdo, případně jen správce systému.

Pomocí velmi výkonných vyhledávacích nástrojů typu Google [1] je na základě klíčových slov pro dané téma ještě snadnější informace na internetu vyhle-

dávat. Vyhledávací stroje mechanicky, bez jakékoli cenzury či třídění, prohlížejí a pro sebe indexují prostor internetu. Výsledek hledání je pak uživateli předkládán v podobě seznamu stránek řazených na základě relevance, tj. průniku vybraných kritérií kontrolovaných provozovatelem vyhledávacího stroje. I přesto, že se vyhledávací stroje často honosí tím, že dávají odpovědi na otázky, spíše než odpovědi se uživateli dostává na vrchních místech seznamu pouze informací o zdrojích nejlépe se shodujících s uživatelským hledáním. Kritéria řazení výsledků vyhledávání ovšem zcela pomíjí jakékoli hodnocení správnosti obsahu, filtrování nepravdivého a většinou ani nezohledňují certifikaci přidělenou kvalitním zdrojům informací.

Jistým omezením pro svobodu publikování informací jsou jen technické znalosti těch, kdo publikují. Omezením však zdaleka nejsou jejich věcné znalosti a kompetentnost vyjadřovat se k dané problematice. Výsledkem je situace, kdy má uživatel hledající na internetu k dispozici řadu často protichůdných odpovědí na jeho hledání a jen velmi omezené možnosti, jak ve výsledcích rozpoznat kvalitní informace od nekvalitních, zkreslených (například reklamou) nebo dokonce zavádějících a klamných. Zvláště významný je tento problém právě v oblasti medicínských informací, které se vyznačují svojí delicate a důležitostí. Jednoduchá dostupnost obrovského množství informací ve velmi variabilní kvalitě (počínaje metaanalýzami nebo systematickými přehledy a popularizačními články konče) pro takto klíčovou oblast lidského života, přináší problémy v řadě ohledů. Správná informace může uživateli přinést řadu pozitivních efektů a v globálním měřítku

i řadu ekonomických úspor pro zdravotnictví jako celek. Naopak spoléhání se na chybné či přímo klamavé informace může mít zcela opačný dopad a způsobit tak nepříjemné následky nebo i trvalé nevratné poškození zdraví u dotčených osob (případně zbytečné dodatečné náklady v celém zdravotnickém systému). Pravdivost dohledaných informací je tedy zcela na prvním místě seznamu kritérií kvality, která se nabízejí.

Dalším projevem snadné dostupnosti stále většího a většího množství informačních zdrojů pro širokou laickou i odbornou veřejnost, je zahlcení velkým množstvím informací. Tento jev se může týkat jak běžných lidí, tak i praktických lékařů. Výsledkem může být například přehlédnutí velmi podstatné informace pro daný případ v záplavě mnoha dalších nepodstatných dat. Často jsou zmiňovány i teoretické problémy komplikovanější komunikace s pacientem, který je předem vybaven nevhodnou nebo chybnou informací.

Bohužel, jediná v současné době reálná možnost, jak zjistit kvalitu existujících medicínských textů, je manuální verifikace odborníkem v oboru. Odborníci, zabývající se tematikou hodnocení kvality medicínských informačních zdrojů na internetu, většinou tyto zdroje posuzují komplexně a vedle vlastního obsahu textů zkoumají také řadu dalších formálních kritérií jako například kvalitu a přehlednost prezentace, přítomnost kontaktních údajů nebo shodu s aktuálními standardy tvorby webových stránek [2]. Na rozdíl od hodnocení obsahu, pro mnohé z těchto dalších formálních aspektů již existují a jsou používány nástroje umožňující jejich poloautomatické hodnocení [3]. Příkladem takového systému podporujícího expertní rozhodování při hodnocení kvality je nástroj AQUA [4] vyvinutý v rámci projektu MedIEQ [5]. Poloautomatické hodnocení obsahu by mohlo být dalším významným vylepšením podobných nástrojů a zefektivněním práce expertů. Prověřené dokumenty mohou být následně označeny známkou či certifikátem kvality (například HON [6] nebo WMA [7]), nebo katalogizovány do kvalitativních či tematických kategorií.

Druhou možnou cestou, jak zajistit vysokou obsahovou úroveň, je, že sám expert sestaví text o dané problematice. Takovéto texty sepsané přímo odborníky v oboru bývají pak publikovány pod záštitou některé renomované lékařské společnosti, což jim dává záruku určité kvality. Nehledě na fakt, že takovýto postup je velmi nákladný, časově zdouhavý a ve velkém měřítku tedy v podstatě nepoužitelný, hlavní problém i nadále přetrvává. I takto prověřené dokumenty totiž mohou zapadnout nerozpoznány v záplavě dalších dostupných textů. Hlavní výzvy pro informatiku v oblasti kvality medicínských informací dnes tedy spočívají ve dvou směrech. Jednak v možnosti jasného definování jediné konsenzuální verze pravdy, založené na nejaktuálnějších vědeckých poznatcích v daném oboru a za druhé na možnosti tento etalon snadno dohledat, porovnávat k němu a odkazovat se na něj při posuzování obsahové kvality všech dalších textů.

Vzhledem k decentralizovanému způsobu vzniku nových vědeckých poznatků, k mnoha vyskytujícím se národním specifikům v systémech zdravotnictví a k existenci řady organizací aspirujících na postavení nejvyšší autority, není reálné očekávat tuto jedinou sdílenou verzi pravdy od nějakého z těchto subjektů. Nejnadějnější se jeví v této souvislosti aktivity spojené s vytvářením tzv. doporučených postupů (DP) [8] pro medicínu, které jsou systematicky připravovány a aktualizovány týmy expertů. Ty jsou pak publikovány pod hlavičkou renomovaných lékařských společností, zdravotnických organizací [9] nebo agentur specializujících se na publikování doporučených postupů [10]. Doporučení jsou sestavována s využitím principů medicíny založené na důkazech (MZD), jejímž základem je hierarchicky uspořádaná struktura důkazů respektive vědeckých prací. Cílem je primárně uplatňovat dostupné důkazy položené co nejvýše v hierarchii jejich síly a významnosti. Nejvýše stojí meta-analýzy a systematické přehledy. DP kompletně pokrývají problematiku léčby dané choroby z pohledu diagnózy, průběhu, léčebných postupů, jejich vzájemné zastupitelnosti, aplikovatelnosti v konkrétních podmínkách a dokonce hodnotí například i relativní

přínos jednotlivých metod ve srovnání s jejich nákladností respektive s obtížemi působícími pacientům. Důležitou vlastností DP je to, že jsou velmi dobře strukturovány. V současné době již existují postupy zabývající se jejich formalizací a převáděním do zcela strukturované elektronické podoby [11]. Ta pak může být implementována například v nemocničních systémech v kombinaci s elektronickými zdravotními záznamy pacientů, nebo právě pro hodnocení informační kvality textových dokumentů na internetu.

Informační kvalita je definována jako hodnota, kterou daná informace přináší jejímu uživateli. Z toho vyplývá, že velmi významnou roli při pohledu na informační kvalitu hraje její subjektivita. Samotná informační kvalita může být nahlížena ze čtyř různých směrů (či dimenzí). První skupinu tvoří vlastnosti přímo se vztahující k vlastní podstatě informace, kde příkladem může být přesnost, objektivita nebo důvěryhodnost informace. Druhou dimenzi tvoří vlastnosti zasazující informaci do kontextu dalších dostupných informací (tj. například úplnost, včasnost, relevantnost nebo přidaná hodnota informace). Třetí dimenze je zastoupena vlastnostmi vyjadřujícími přijetí čtenářem, tedy vlastnostmi jako pochopitelnost, snadnost porozumění, zhuštěnost sdělení nebo logická konzistentnost. Poslední aspekt informační kvality je spojen s dostupností informace pro uživatele (tj. například snadnost její získání respektive aktualizace nebo zabezpečení přístupu). Aby bylo informační kvalitu možné nějakým způsobem měřit a hodnotit, je potřeba z výše popsaných možností vybrat takové vlastnosti, které jsou co možná nejméně subjektivní a na jejich základě pak vytvořit metriky informační kvality. Tato práce se svým tématem dotýká zejména objektivních vlastností informační kvality, jako jsou úplnost pokrytí či nepokrytí daného tématu, použití odborné terminologie, přesnosti, důvěryhodnosti, ověřitelnosti a přístupnosti informací. Subjektivnost předávané informace je vnímána autory textů, kteří se následně snaží uzpůsobovat své příspěvky konkrétním skupinám čtenářů.

V oblasti medicínských textů na internetu je možno rozlišit texty určené jednak laické veřejnosti (pacienti nebo děti) a dále texty určené profesionálům (tj. např. lékařům nebo vědeckým pracovníkům v medicíně). Texty, určené pro jednotlivé z těchto skupin, se ve výsledku velmi zásadně liší právě v řadě ohledů řazených mezi ty subjektivní. Například použití přesné odborné terminologie zvyšuje přesnost vyjádření a je profesionály v oboru velmi ceněno. U laiků může ale na druhé straně zásadně snížit snadnost pochopení textu. U subjektivních vlastností, je ovlivnění čtenářskou kategorií zcela zjevné. Tato kategorizace má ale určitý vliv i na vlastnosti objektivnějšího charakteru a její zohlednění při hodnocení informační kvality se zdá být tedy nanejvýš vhodné.

2. Cíle

Cílem této práce je předložit jednoduchý postup hodnocení kvality obsahu webových stránek (WS). Doporučené postupy (DP) jsou využity jako standard kvality a řízené medicínské slovníky zde slouží jako prostředek, pomocí kterého jsou obě skupiny dokumentů porovnávány. Kvalita WS bude stanovována jednak na základě obecnější shody obsahu (tedy vyskytujících se konceptů a témat) a dále na základě podobnosti konkrétní použité odborné terminologie s koncepty a terminologií vyskytující se v DP. Dílčím cílem je navrhnout a vyhodnotit vhodné metody agregace terminologie DP tak, aby bylo možno testované WS porovnávat vůči jedinému měřítku i v případě, že je dostupných DP více. Posledním cílem práce je vyhodnocení použitelnosti tohoto postupu v procesu poloautomatického hodnocení kvality. Důraz bude kladen na popis silných i slabých míst přístupu a zhodnocení jejich dopadu na možnosti praktické aplikace.

3. Metody

K vybranému medicínskému tématu byly nalezeny jednak existující doporučené postupy (DP) a dále zhruba sto webových stránek (WS). V obou skupinách bylo provedeno dohledání odborné terminologie a výsledky získané z DP byly použity jako měřítko pro hodnocení obsahu WS.

Téma „lung cancer screening“, na kterém byl postup testován, bylo vybráno tak, aby

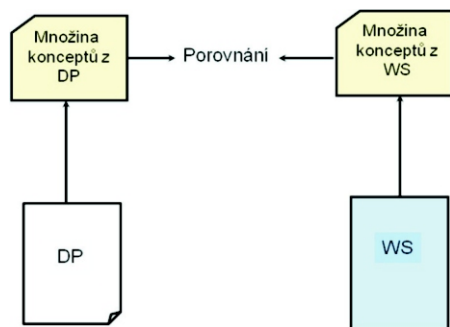
bylo jasně tematicky vymezené, dále aby pro něj existovaly použitelné DP a aby bylo přiměřeně přístupné i laické veřejnosti. Celá práce byla prováděna jen na anglicky psaných dokumentech (DP, WS). Prohledáním existujících databází a katalogů doporučených postupů byly pro metodu vybrány nejhodnější DP. Vedle DP bylo provedeno i hledání existujících WS s příslušnou tematikou. Dokumenty použité jako korpus byly získány jednoduše jako prvních zhruba sto odkazů vrácených vyhledávačem Google [1] pro hledání řetězce „lung cancer screening“. Následně byly lokálně uloženy nástrojem ScrapBook [12]. Soubor dokumentů byl ručně dorovnán na rovných sto dokumentů po vyřazení nefunkčních nebo nestahitelných stránek. Stejně tak byly vyřazeny odkazy na vybrané DP, z nichž se některé také umístily mezi prvními 100 výsledky vrácenými vyhledávačem. Tyto DP byly vyřazeny, aby nedocházelo ke zkreslení při porovnávání sama se sebou. Na základě odhadu cílové skupiny čtenářů (diskutováno v úvodu), byly dále dokumenty rozděleny do několika kategorií.

Texty doporučených postupů byly anotovány pomocí nástrojů postavených nad UMLS metatezaurem [13], [14]. Původně bylo mapování prováděno pomocí MMTx (MetaMap Transfer) anotátoru [15], ale později se ukázalo jako dostatečné a dokonce vhodnější využití nástroje Interactive MetaMap [16]. Oba nástroje byly vyvinuty NLM [17], což je organizace, která také vyvíjí samotné UMLS. V první fázi byly celé texty zpracovány jedním z nástrojů, přičemž výstupem byl seznam nalezených termínů. Ke každému z těchto termínů byly následně dohledány odpovídající koncepty pomocí SQL dotazování nad lokálně uloženou databází UMLS (MeSH, ICD-10 apod.). Výsledkem mapování byl seznam respektive hierarchický strom odborných výrazů a jejich rodičovských konceptů, tj. množiny popisující obsah DP. Podobné mapování na UMLS, bylo provedeno i pro všechny WS. Produkty mapování pro obě skupiny pak byly vzájemně srovnány.

Terminologie použitá ve WS ale často neodpovídá terminologii použité v DP a dokonce ani terminologii obsažené v UMLS. V UMLS jsou sice k oficiálním termínům uváděna synonyma, ale většinou

jde zase jen o odborné termíny nebo jiné názvy, používané v dalších databázích metatezauru. Jako synonyma například nejsou často k dispozici hovorové, méně přesné nebo zkrácené názvy jak nemocí, metod nebo třeba přístrojů. Všechna tato synonyma se nezbytně vyskytují v textech určených laickému čtenáři. Například koncept v UMLS vedený jako „Tomography, Spiral Computed“ je v realitě zastupován celou řadou synonym, zkrácených nebo neúplných názvů a zkratk jako např. „CAT Scan, Spiral“, „Computed Tomography, Spiral“, „Computer-Assisted Tomography, Spiral“, „Computerized Tomography, Spiral“, „CT Scan, Spiral“ a podobně. V testovaných WS stránkách byla tato metoda, ale velmi často odkazována jen výrazem „scan“, což vedlo k minutí se s terminologií UMLS nebo naopak k nejasnému či nesmyslnému mapování. Z pohledu obsahu dokumentu se tedy obě sady často lišily i v případě, že se jednalo o čistě syntaktický (názvoslovný), nikoli sémantický rozdíl. Postup metody pro finální porovnání musel být proto rozšířen tak, aby byl schopný brát v úvahu i termíny vyskytující se pouze ve WS (tj. chybějící jak v DP tak v UMLS).

Množina konceptů pro vybrané téma a jim odpovídající synonymické termíny získané z prvního kola mapování, byla tedy následně databázově uložena mimo struktury UMLS. Nalezenými termíny oannotovány testované WS (nebo alespoň několik z nich) byly pak manuálně kontrolovány, zda je v nich veškerá terminologie plně pokryta termíny z UMLS. Pokud tomu tak nebylo, byly chybějící termíny postupně doplněny do uloženého seznamu terminologie. Doplnění proběhlo jen pro chybějící výrazy, zcela jasně zařaditelné pod vybrané koncepty (typicky šlo o různé varianty existujících synonym). Tento ruční krok umožnil následně úplnější mapování konceptů pro všechny dokumenty a zkvalitnil jejich vzájemné porovnání.



Obr. 1. Schéma procesu porovnání WS a DP.

Vzhledem k tomu, že obohacený seznam terminologie již nebyl uložen v rámci UMLS, závěrečná anotace WS a DP byla provedena nástrojem Super Text Search [18], který umožňuje plnotextové prohledávání seznamu dokumentů. Pro takto anotované soubory byly zaznamenány vyskytující se distinktní odborné termíny a jejich celkové počty výskytů v každém z dokumentů. Na základě vztahů synonymy byly z konkrétních termínů vyvozeny distinktní koncepty a znovu spočítány jejich výskyty.

Ze zaznamenaných výsledků byla nakonec počítána kosinová podobnost mezi DP a WS. Jak bylo zmíněno výše, tak vhodné DP byly nalezeny a použity tři (dále značené a_0 , a_1 , a_2). Bylo tedy třeba stanovit, proti jakému standardu bude nejvhodnější WS porovnávat. V této souvislosti byly navrženy čtyři různé způsoby agregace množin na DP namapovaných termínů respektive konceptů. Jednalo se o „průnik“ množin termínů (značeno $a_1 \cap a_2 \cap a_0$) – tj. jen distinktní termíny, které byly pro všechny DP společné, dále pak o „sjednocení“ ($a_1 \cup a_2 \cup a_0$), tj. distinktně všechny termíny nalézající se minimálně v jednom DP, dále o „sumu“ ($\text{sum}(a_1, a_2, a_0)$), tj. jednoduše sloučené seznamy nalezených termínů včetně počtů jejich výskytů a o konečně „váženou sumu“ ($\text{nsum}(a_1, a_2, a_0)$), která navíc při spojování množin dokumentů zohledňovala rozsah každého DP (celkový počet všech nalezených výskytů jakéhokoli termínu). Vyhodnocení vhodnosti agregací bylo realizováno pomocí aplikace kosinové podobnosti křížem mezi všemi produkty agregací a jednotlivých DP. Toto srovnání vzájemné podobnosti bylo provedeno za prvé s využitím úplných

množin všech nalezených termínů, za druhé s využitím booleovsky transformovaných množin termínů (tj. jen distinktní výskyty termínů v dokumentech) a konečně s využitím mapovaných konceptů.

Porovnání kosinové podobnosti všech WS bylo prováděno proti všem verzím agregovaných standardů i proti všem doporučeným postupům. Porovnání znovu proběhlo na základě využití distinktních i celkových výskytů odborné terminologie. Výsledkem je údaj udávající „terminologickou podobnost“. Analogicky byla počítána i podobnost na základě distinktních (případně absolutních počtů) zastoupených konceptů jako „podobnost obsahová“. Průměrná kosinová podobnost WS se standardy byla počítána za jednotlivé kategorie dokumentů.

4. Výsledky

Metoda byla aplikována na medicínské téma „screening for lung cancer“. Testovací webové stránky byly získány vyhledáváním řetězce „lung cancer screening“ ve vyhledávači Google. Hledání řetězce Googlem vrátilo přibližně 2 miliony záznamů. Pro porovnání bylo provedeno i podobné hledání vyhledávači Yahoo! [19] (29 milionů záznamů) a českým vyhledávačem Seznam [20] (jen 120 záznamů). Jako korpus testovacích stránek bylo ale použito prvních 100 nejrelevantnějších odkazů (po očištění nepoužitelných odkazů tj. nefunkčních linků a DP) vrácených Googlem. Vyřazené doporučené postupy, se umístily ve druhé polovině vrchních 100 výsledků vrácených Googlem. Webové stránky byly následně lokálně uloženy nástrojem ScrapBook.

V dostupných informačních zdrojích, tedy existujících databázích nebo katalozích doporučených postupů, a dále volně na internetu bylo nalezeno několik dokumentů DP odpovídajících vybranému tématu. Z nalezených DP byly jako nejvhodnější vybrány tři. Vyřazené DP byly buď národně či jazykově specifické, obsahem neúplně pokrývaly téma nebo byly naopak mnohem obecnější. V jednom případě se jednalo o starší verzi jednoho ze tří použitých doporučení.

Testované WS byly ručně zaklasifikovány do několika kategorií v závislosti na jejich charakteru a na cílové skupině uživatelů

textu. Pro námi vybrané téma byly dohledány stránky spadající do pěti kategorií. První skupinu tvoří stránky určené pro odbornou lékařskou veřejnost, které byly zastoupeny 23 procenty WS. Druhou významnou skupinou jsou odborné publikace, které jsme dále rozdělili na ty dostupné v plném znění nebo alespoň ve formě abstraktu nebo souhrnu (23 %) a na ty, pro něž byl dostupný pouze název nebo jen velmi krátký souhrn (5 %). Důležitou skupinou dokumentů jsou WS určené pro laickou veřejnost – tedy jednak pro pacienty (21 %) a dále pro děti (0 %). Ačkoli jsou články určené dětským čtenářům pro jiná medicínská témata běžné (většinou mají vzdělávací nebo preventivní charakter), pro vybrané téma nebyly v korpusu WS vůbec zastoupeny (pravděpodobně pro přílišnou specifičnost a odbornost tématu). Poslední kategorie byla vytvořena uměle pro texty neurčené pro žádnou konkrétní skupinu uživatelů (28 %). V tomto případě jde zejména o obecné zprávy, prohlášení, novinové články a podobně.

Doporučené postupy byly oantovány pomocí MMTx respektive Interactive MetaMap. Výsledkem procesu byly texty s namapovanými odbornými termíny UMLS, přičemž přiřazené termíny současně odpovídají odborným konceptům, které pak zastupují obsah textu. Tímto postupem bylo vyvozeno 15 konceptů významných pro vybranou tematickou oblast. Na základě takto omezené množiny odborné terminologie byly podobně anotovány i testovací WS. Výsledné anotované texty obou skupin, jak DP, tak WS, byly nakonec manuálně analyzovány s cílem dohledat nedostatky v MetaMap mapování na UMLS. Bylo zjištěno, že u DP je termínů, které by nebyly zachyceny anotátorem, jen velmi málo. Naopak u skupiny WS byla úspěšnost mapování odhadnuta jen na zhruba 60 %. Obě tato zjištění plně odpovídají očekávání a jsou

Tab. 1. Počty WS korpusu po kategoriích.

typ	popis	počet
m	WS určené lékařům	23
p	WS určené pacientům	21
ch	WS určené dětem	0
g	obecné, novinové články, zprávy, další nezařaditelné	28
mo	vědecké publikace (v plném znění nebo alespoň abstrakt)	23
mo/x	vědecké publikace (omezený přístup, obvykle jen název)	5

zjevně důsledkem faktu, že UMLS je určen zejména pro práci s texty psanými odbornou terminologií.

Všech 15 namapovaných konceptů a veškerá jejich synonyma obsažená v UMLS byla uložena do nové hierarchické databázové struktury. Originální seznam synonym byl ručně rozšířen o chybějící výrazy zjištěné během revize anotovaných WS. Takto rozšířeným seznamem terminologie byly následně znovu anotovány všechny dokumenty a to s již mnohem vyšší úspěšností.

Pro porovnání byly v našem případě k dispozici tři různé DP. Aby porovnání podobnosti nemuselo v budoucnosti probíhat ke každému dostupnému DP zvlášť, bylo jedním z cílů práce navrhnout a otestovat postupy, které by umožnily množiny anotovaných termínů pro jednotlivé DP agregovat do jedné společné. Jednotlivé verze agregovaných množin i množin zastupujících DP byly vzájemně porovnány kosinovou podobností.

Jako metody agregací, pro něž se výsledek porovnání v průměru nejlépe shodoval s původními doporučenými postupy, byly vyhodnoceny $\text{sum}(a1,a2,a0)$ a $\text{nsum}(a1,a2,a0)$. Podobnost u těchto dvou metod se v závislosti na použité variantě pohybovala v rozmezí od 0,90 do 0,99. Stupeň vzájemné podobnosti množin termínů zastupujících tři DP se pohybovala na podobné úrovni, tedy v hodnotách blízkých se 1. Agregace, založené na průniku a sjednocení, vyšly ze srovnání podobnosti proti DP jako relativně méně vhodné, přičemž nejnižší průměrné hodnoty podobnosti byly zaznamenány pro průnik.

Analýza vzájemné podobnosti byla provedena na třech úrovních detailu. První způsob porovnával veškeré výskyty odborných termínů a zohledňoval počet jejich výskytů v textu jako váhu daného termínu (v tabulkách označováno jako „similarity (terms)“). Druhý způsob taktéž pracoval s termíny, ale porovnání bylo zjednodušeno jen na jejich distinktní výskyt v textu („similarity(distinct terms)“). Poslední způsob využíval k porovnání podobnosti koncepty namapované přes odborné výrazy nalezené v textu („similarity (concepts)“). Z tohoto pohledu bylo

nejvyšších průměrných hodnot podobnosti dosahováno u konceptů, dále u distinktních termínů a nejnižších při využití absolutních počtů výskytů termínů. Souhrnné výsledky křížového porovnání DP a jejich agregací jsou uvedeny v následujících tabulkách.

Tab. 2a. Křížové podobnosti mezi množinami termínů zastupujícími jednotlivé dokumenty.

Označení „a0, a1, a2“ odpovídá třem DP, „ $a1 \cap a2 \cap a0$ “ odpovídá průniku, „ $a1 \cup a2 \cup a0$ “ značí sjednocení, „ $\text{sum}(a1,a2,a0)$ “ značí sumu a „ $\text{nsum}(a1,a2,a0)$ “ značí váženou sumu. Maximální 100% podobnost (tj. identita) je reprezentována hodnotou 1. Hodnota 0 znamená absolutní nepodobnost obou množin..

dokument	similarity (terms)						
			$\text{nsum}(a1,a2,a0)$	$\text{sum}(a1,a2,a0)$	a0	a1	a2
$a1 \cap a2 \cap a0$	1,00	0,66	0,70	0,71	0,73	0,63	0,69
$a1 \cup a2 \cup a0$							
	0,66	1,00	0,58	0,61	0,64	0,44	0,64
$\text{nsum}(a1,a2,a0)$	0,70	0,58	1,00	1,00	0,99	0,95	0,95
$\text{sum}(a1,a2,a0)$	0,71	0,61	1,00	1,00	1,00	0,92	0,97
a0	0,73	0,64	0,99	1,00	1,00	0,89	0,98
a1	0,63	0,44	0,95	0,92	0,89	1,00	0,81
a2	0,69	0,64	0,95	0,97	0,98	0,81	1,00

Tab. 2b. Křížové podobnosti mezi množinami konceptů zastupujícími jednotlivé dokumenty.

Označení „a0, a1, a2“ odpovídá třem DP, „ $a1 \cap a2 \cap a0$ “ odpovídá průniku, „ $a1 \cup a2 \cup a0$ “ značí sjednocení, „ $\text{sum}(a1,a2,a0)$ “ značí sumu a „ $\text{nsum}(a1,a2,a0)$ “ značí váženou sumu. Maximální 100% podobnost (tj. identita) je reprezentována hodnotou 1. hodnota 0 znamená absolutní nepodobnost obou množin.

dokument	similarity (concepts)						
			$\text{nsum}(a1,a2,a0)$	$\text{sum}(a1,a2,a0)$	a0	a1	a2
$a1 \cap a2 \cap a0$	1,00	0,91	x	0,91	0,91	1,00	0,91
$a1 \cup a2 \cup a0$							
	0,91	1,00	x	1,00	1,00	0,91	1,00
$\text{nsum}(a1,a2,a0)$	x	x	x	x	x	x	x
$\text{sum}(a1,a2,a0)$	0,91	1,00	x	1,00	1,00	0,91	1,00
a0	0,91	1,00	x	1,00	1,00	0,91	1,00
a1	1,00	0,91	x	0,91	0,91	1,00	0,91
a2	0,91	1,00	x	1,00	1,00	0,91	1,00

Tab. 2c. Křížové podobnosti mezi množinami distinktních termínů zastupujícími jednotlivé dokumenty. Označení „a0, a1, a2“ odpovídá třem DP, „ $a1 \cap a2 \cap a0$ “ odpovídá průniku,

„ $a1 \cup a2 \cup a0$ “ značí sjednocení, „ $\text{sum}(a1,a2,a0)$ “ značí sumu a „ $\text{nsum}(a1,a2,a0)$ “ značí váženou sumu. Maximální 100% podobnost (tj. identita) je reprezentována hodnotou 1. hodnota 0 znamená absolutní nepodobnost obou množin.

dokument	similarity (distinct terms)						
			$\text{nsum}(a1,a2,a0)$	$\text{sum}(a1,a2,a0)$	a0	a1	a2
$a1 \cap a2 \cap a0$	1,00	0,66	x	0,82	0,71	0,88	0,71
$a1 \cup a2 \cup a0$							
	0,66	1,00	x	0,96	0,94	0,75	0,94
$\text{nsum}(a1,a2,a0)$	x	x	x	x	x	x	x
$\text{sum}(a1,a2,a0)$	0,82	0,96	x	1,00	0,94	0,86	0,94
a0	0,71	0,94	x	0,94	1,00	0,71	0,86
a1	0,88	0,75	x	0,86	0,71	1,00	0,71
a2	0,71	0,94	x	0,94	0,86	0,71	1,00

Analogicky tomu, jak byly vzájemně porovnány množiny reprezentující obsah DP, byly proti těmto množinám kosinovou podobností porovnány i obsahy všech WS. Toto porovnání znovu probíhalo na třech různých úrovních detailu tedy na úrovni počtů termínů, na úrovni distinktních termínů a na úrovni konceptů.

Průměrná podobnost přes všechny WS, přes všechny DP (a jejich agregace) a přes všechny tři typy detailu dosáhla 0,69. Obecně nejnižších podobností bylo dosaženo při analýze na úrovni distinktních termínů (průměrná hodnota 0,56 v porovnání s 0,745 pro koncepty a 0,75 pro termíny). Podobně jako při křížovém porovnání standardů, vyšly nejvyšší průměrné podobnosti korpusu WS k jednotlivým DP a jejich agregacím znovu pro agregace $\text{sum}(a1,a2,a0)$ respektive $\text{nsum}(a1,a2,a0)$. O něco nižší hodnoty vyšly pro neagregované DP a nejnižší hodnoty pro sjednocení a průnik.

Podobnosti vyčíslené přes jednotlivé kategorie dokumentů se odchylovaly od celkového průměru v průměrném rozsahu 6,9 % pro termíny a koncepty a v rozsahu 9 % pro distinktní termíny. Obecně nejvyšší shoda WS a DP byla nalezena pro kategorii „mo“ tj. vědeckých publikací (průměrně 0,78) a naopak nejnižší pro kategorie „g“ (obecné texty) a „mo/x“ (neúplné vědecké publikace).

Mezi porovnáním pomocí termínů a konceptů je nejen metodický ale i zásadní významový rozdíl. Při porovnání termínů jde o porovnání terminologické podobnosti, kdežto při analýze podobnosti na základě konceptů jde skutečně o porovnání obsahové. Průměrná terminologická podobnost pro korpus dokumentů vyšla 0,76 a průměrná obsahová podobnost pak 0,85.

5. Diskuze

Vzhledem k tomu, že se v této práci jednalo o první pokusy porovnání DP a WS, bylo potřeba pečlivě vybrat téma, na kterém by bylo možné celou metodiku porovnání demonstrovat a ověřit ji. Téma muselo být vybráno tak, aby pro něj existovaly vhodné DP (tj. aby dané téma bylo DP pokryto kompletně a aby naopak netvořilo jen podmnožinu DP). K vybranému tématu bylo nakonec k dispozici dokonce několik DP, což umožnilo vyvinout a zhodnotit některé potenciálně využitelné způsoby agregace množin termínů zastupujících obsah tak, aby byl získán jediný testovací standard pro hodnocení obsahu WS. Při výběru tématu bylo rovněž zohledňováno, aby v prvních 100 WS byly co nejrovnoměrněji zastoupeny stránky pro rozdílné skupiny čtenářů. Ze všech čtenářských skupin, které teoreticky přicházely v úvahu, nebyly v korpusu dokumentů zastoupeny jen stránky určené dětem.

Představa možnosti zobecnit tento postup na jakékoli medicínské téma je ovšem spojena s řadou komplikací. Prvním

problémem je fakt, že postup předpokládá v budoucnosti pro DP systematické pokrytí celé domény medicíny. K tomu je ovšem dnešní praxe hodně vzdálena jak rozsahem pokrytí, tak organizovaností vzniku DP. DP vznikají velmi distribuovaně, nepravidelně a tematicky pokrývají v podstatě jen některé z nejdůležitějších oblastí. DP jsou navíc jazykově omezené vždy na jednu konkrétní lingvistickou oblast, což představuje další překážku jak jejich většímu rozšíření, tak při jejich konkrétní aplikaci.

Na aplikovatelnost metody má také velký vliv pokrytí medicínské terminologie UMLS metatezaurem. I přesto, že je UMLS pravidelně aktualizováno, stále rozšiřováno o nové a nové zdroje a ve výsledku má velmi kvalitní pokrytí konceptů, řada dílčích termínů v UMLS chybí. Pro každý z konceptů je k dispozici řada synonym a je tedy možno využít tento slovník k mapování terminologie na slova a slovní spojení nalezená v odborných textech. Primárním cílem UMLS je ale být slovníkem správné terminologie. Velmi často proto chybí hovorové, nesprávné nebo zastaralé názvy, což má za následek, že mapování často selhává u textů psaných běžným jazykem, který se terminologicky značně liší od jazyka odborného. To se ukázalo i v této práci, kdy mapování na profesionální terminologii psané DP bylo téměř stoprocentně úspěšné, kdežto mapování na WS psané běžným jazykem dosáhlo úspěšnosti jen kolem 60 %.

Aby bylo možné dále testovat úroveň konceptuální shody, rozšířili jsme ad-hoc pro testovaný postup seznamy synonym u konceptů týkajících se vybraného tématu na základě předanotovaných WS. Během ručního posouzení WS se ukázalo, že na prvních 10-15 dokumentech byla nalezena převážná většina chybějících termínů a zbylé WS byly v podstatě kontrolovány už jen pro úplnost. Na základě rozšířeného seznamu terminologie pak už byly velmi úspěšně anotovány jak DP, tak i WS. Takovýto ruční zásah, ale samozřejmě není obecně aplikovatelný hromadně pro všechna medicínská témata a je zjevnou slabinou automatizovatelnosti a obecnosti využití celého postupu.

Vedle problémů spojených s úplností UMLS, tj. s výskytem jednoho konceptu pod různými názvy neboli synonymie, jsou značnou překážkou spolehlivého mapování i další vlastnosti přirozeného jazyka [21]. Asi tím nejvýznamnějším problémem pro počítačové zpracování textů je homonymie [22]. Aby bylo například možné přesně určit, o který z významů slova se v dané souvislosti jedná, je většinou nutné posoudit okolní kontext slova a skutečně pochopit význam textu.

Pokud necháme stranou výše diskutované dílčí problémy procesu, je možné si představit využití této metodiky v procesu poloautomatického hodnocení kvality webových dokumentů následovně. Prvním krokem takového procesu by bylo stanovení tématu testovaného textu například s využitím zmíněného nástroje AQUA [4]. Nalezené téma by sloužilo jednak pro dohledání všech relevantních doporučených postupů k danému tématu a za druhé by byla pomocí UMLS nástrojů vyhledána veškerá relevantní terminologie. Následně by proběhlo vyhodnocení podobnosti textu s doporučenými postupy. Výsledná míra podobnosti by pak mohla sloužit jako doplňující podklad pro expertní rozhodování, přičemž klasifikace dokumentu do jedné z kategorií by pomáhala tuto hodnotu podobnosti lépe interpretovat, případně použít jen pro dokumenty spadající do vybraných kategorií.

Tab. 3. Průměrné kosinové podobnosti kategorií WS proti DP a jejich agregacím. Na svislé ose jsou uvedeny kategorie WS a na ose horizontální DP a jejich agregace při porovnání podobnosti na třech úrovních detailu (termíny, distinktní termíny a koncepty). Červeně jsou znázorněny maximální a modře minimální hodnoty přes jednotlivé kategorie a přes všechna data.

kategorie / dokument	similarity (terms)						similarity (concepts)						similarity (distinct terms)					
	a1	Ua0	nsim(a1, a2, a0)	sum(a1, a2, a0)	a0	a1	a2	a1	Ua0	nsim(a1, a2, a0)	sum(a1, a2, a0)	a0	a1	a2	a1	Ua0	nsim(a1, a2, a0)	sum(a1, a2, a0)
g	0,53	0,43	0,83	0,82	0,81	0,78	0,80	0,75	0,75	-	0,75	0,75	0,75	0,75	0,50	0,53	-	0,56
m	0,59	0,50	0,86	0,87	0,87	0,78	0,87	0,75	0,77	-	0,77	0,77	0,75	0,77	0,51	0,63	-	0,64
ch	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
mo	0,60	0,53	0,88	0,89	0,89	0,80	0,88	0,72	0,79	-	0,79	0,79	0,72	0,79	0,52	0,67	-	0,67
mo/x	0,57	0,48	0,85	0,85	0,86	0,76	0,85	0,64	0,73	-	0,73	0,73	0,64	0,73	0,50	0,53	-	0,57
p	0,56	0,45	0,84	0,83	0,82	0,79	0,81	0,75	0,73	-	0,73	0,73	0,75	0,73	0,52	0,57	-	0,60
all	0,57	0,48	0,85	0,85	0,85	0,79	0,84	0,74	0,76	-	0,76	0,76	0,74	0,76	0,51	0,59	-	0,61

5. Závěry

Dosavadní výsledky jsou prvním krokem směrem k automatickému hodnocení obsahu medicínských webových zdrojů s využitím DP, jakožto standardu kvality. Hlavním cílem práce bylo navrhnout postup, vyhodnotit jeho praktickou použitelnost a stanovit další směry výzkumu. V této fázi byly provedeny pokusy na jednom konkrétním pečlivě vybraném tématu, pro které existovaly jak vhodné DP, tak i vhodné WS. Téma bylo modelově zpracováno pro anglicky psané texty.

Pro získání lepší představy o možnostech zobecnění postupu na jakákoli další medicínská témata, by bylo vhodné provést další pokusy s náhodně vybranými tématy a pokusit se zautomatizovat manuální kroky, které postup obsahuje. Z výsledků vyplývá, že podobnost mezi dokumenty nalezenými na internetu a DP závisí mimo jiné i na kategorii čtenářů, pro kterou je text určen. Podobnosti vyčíslené přes jednotlivé kategorie dokumentů se odchylovaly od celkového průměru 6,9 % (pro termíny a koncepty) a v rozsahu 9 % pro distinktní termíny. Obecně nejvyšší shoda WS a DP byla nalezena pro kategorii vědeckých publikací (průměrně 0,78) a naopak nejnižší pro kategorie „g“ (obecné texty) a „mo/x“ (neúplné vědecké publikace). Tato kategorizace WS byla v této práci ovšem provedena zcela ručně. Nadějnou cestou jak zautomatizovat tuto kategorizaci WS by bylo využít existující funkčnosti vícejazyčného nástroje AQUA [2], který byl vyvinut pro poloautomatické zpracování WS. Na druhé straně využití zjištění, že samotné zaklasifikování dokumentu do jedné z námi použitých kategorií už nepřímo vypovídá o jeho shodě s doporučeným postupem a nepřímo tedy i informační kvalitě textu, by mohlo sloužit jako dílčí podklad pro expertní rozhodování uživatelů tohoto systému.

Stejně tak byly manuálně vyhledány DP a vybrány nejvhodnější z nich. Situace v tomto ohledu se v budoucnosti může zlepšit s tím, jak se katalogy DP na internetu vyvíjí a rozšiřují [7], [8]. Pro zpracování DP byly navrženy postupy umožňující automatickou agregaci několika dostupných DP do jedné sady termínů a vybrán nejvhodnější z nich. K tomuto

etalonu pak mohou být porovnávány testované WS.

Dalším manuálním krokem postupu bylo ruční rozšíření seznamu synonym pro UMLS koncepty nutné pro následné úspěšné anotování dokumentů psaných běžným jazykem. Na druhou stranu bylo ale ukázáno, že pro nalezení chybějících výrazů zcela postačuje projít jen prvních několik předanotovaných WS, řazených vyhledávačem podle jejich relevance.

Množiny konceptů reprezentují obsah dokumentu jen do určité míry. Velkou výzvou, navazující na prosté porovnání množin zastupujících terminologii nebo obsah dokumentů, by bylo porovnání vlastního sdělení textů například pomocí konfliktních tvrzení mezi WS a DP [23].

Poděkování

Částečně podpořeno projektem MŠMT 1M06014 a také grantem SVV-2010-265513.

Literatura

- [1] <http://www.google.com> (accessed September 10, 2010)
- [2] Curro V., Buonomo P.S., Onesimo R., de RP, Vituzzi A., di Tanna G.L., D'Atri A.: A quality evaluation methodology of health web-pages for non-professionals. *Med Inform Internet Med.* 2004;29(2):95-107.
- [3] Wang Y., Liu Z.: Automatic detecting indicators for quality of health information on the Web. *Int J. Med Inform.* 2006;May 31.
- [4] Stamatakis K., Chandrinos K., Karkaletsis V., Mayer M.A., Gonzales D.V., Labsky D.V., Amigó E., Pöllä M.: AQUA, a system assisting labelling experts assess health web resources. 12th Intern. Symposium for Health Information Management Research (iSHIMR 2007), Sheffield, UK. 2007;18-20 July:75-84.
- [5] Mayer M.A., Karkaletsis V., Stamatakis K., Leis A., Villarroel D., Thomeczek C.: "MedIEQ Quality Labelling of Medical Web Content Using Multilingual Information Extraction." *Stud Health Technol Inform.* 2006;121:183-190.
- [6] <http://www.hon.ch> (accessed September 10, 2010)
- [7] Mayer M.A., Leis A., Sarrias R., Ruiz P.: Web Médica Acreditada Guidelines: reliability and quality of health information on Spanish-Language websites. In: Engelbrecht R et al. (ed.). *Connecting Medical Informatics and Bioinformatics. Proc of MIE2005.* 2005;1287-92.
- [8] Field M.J., Lohr K.N. (Eds):. *Guidelines for clinical practice: from development to use.* Institute of Medicine, Washington, D.C: National Academy Press; 1992.
- [9] <http://www.who.int> (accessed September 10, 2010)
- [10] <http://www.guideline.gov> (accessed September 10, 2010)
- [11] Vesely A., Zvarová J., Peleska J., Buchtela D., Anger Z.: Medical guidelines presentation and comparing with Electronic Health Record. *Int J Med Inform.* 2006;Mar-Apr:75(3-4):240-5. Epub 2005; Sep 15.
- [12] <http://amb.vis.ne.jp/mozilla/scrapbook/> (accessed September 10, 2010)
- [13] Lindberg D.A., Humphreys B.L., McCray A.T.: The Unified Medical Language System. *Meth Inform Med.* 1993;32:281-91.
- [14] <http://www.nlm.nih.gov/research/umls> (accessed September 10, 2010)
- [15] <http://mmtx.nlm.nih.gov> (accessed September 10, 2010)
- [16] <http://skr.nlm.nih.gov/interactive/metamap.shtml> (accessed September 10, 2010)
- [17] <http://www.nlm.nih.gov> (accessed September 10, 2010)
- [18] <http://www.galcott.com/ts.htm> (accessed September 10, 2010)
- [19] <http://www.yahoo.com> (accessed September 10, 2010)
- [20] <http://www.seznam.cz/> (accessed September 10, 2010)
- [21] Baud R.H., Ruch P., Gaudinat A., Fabry P., Lovis C., Geissbuhler A.: Coping with the variability of medical terms. *Medinfo.* 2004;11(Pt 1):322-6.
- [22] Rak D., Svatek V., Fidalgo M., Alm O.: Detecting MeSH Keywords and Topics in the Context of Website Quality Assessment. In: *The 1st International Workshop on Describing Medical Web Resources (DRMed 2008)*, held in conjunction with the 21st International Congress of the European Federation for Medical Informatics (MIE 2008), Goteborg, Sweden. 2008: May 27.
- [23] Kaiser K., Miksch S.: Versioning Computer-Interpretable Guidelines: Semi-Automatic Modeling of 'Living Guidelines' Using an Information Extraction Method, *Artificial Intelligence in Medicine (AIIM)*, 2008;55-66.

Kontakt

Mgr. Dušan Rak
U zátíší 545/9
14700 Praha 4–Hodkovičky
e-mail: dušan.rak@seznam.cz

Zrychlení algoritmu pro nalezení optimální šířky jádra v analýze sledu akčních potenciálů

P. Šanda¹

¹Fyziologický ústav AV ČR, v.v.i.

Školitel: Doc. RNDr. Petr Lánský, Csc.

Souhrn

Jednou z důležitých úloh při zkoumání struktury sledu akčních potenciálů je odhad tzv. firing rate funkce. Cílem článku je popsat možné zrychlení algoritmu, který může být pro tento odhad použit.

Protože neexistuje jednoznačná metoda, která firing rate funkci odvodit, bylo navrženo celé spektrum odlišných strategií. Jedna z populárních metod odhadu je konvoluce sledu akčních potenciálů gaussovským jádrem s patřičnou šířkou. Výběr konkrétního jádra a šířky je obvykle diskutabilní a autoři v nedávném článku [1] navrhuji přesný algoritmus pro výpočet optimální šířky pro (nejen) gaussovská jádra. Pro rozsáhlejší množinu vstupních dat je elementární verze algoritmu bohužel neefektivní z hlediska času potřebného pro výpočet. V příspěvku navrhujeme vylepšenou implementaci algoritmu, která je efektivní i pro velká množství vstupních dat.

Na konkrétních výsledcích implementovaného algoritmu bylo demonstrováno dosažené zrychlení, což potvrzuje vhodnost navrhované metody.

Klíčová slova: akční potenciál, sled akčních potenciálů, neurální kódování, firing rate, konvoluce, gaussovské jádro, šířka jádra, Brentova minimalizace, paralelní výpočet, MPI

1. Úvod

Množství prací v početních neurovědách je založeno na předpokladu, že většina informačního toku mezi neurony je realizována pomocí sledu akčních potenciálů. Studium struktury takovéto sekvence vedlo k formulaci řady modelů neuronálního kódu a v mnoha případech jsou navrženy modely schopné úspěšně predikovat vlastnosti experimentálního podnětu, který k sledu akčních potenciálů vedl. Jeden druh úspěšných modelů kóduje důležité vlastnosti stimulu do tzv. firing rate funkce, bohužel se přesný

postup konstrukce této funkce ve většině prací různí a existuje celé spektrum odlišných metod [2].

V tomto příspěvku se zabýváme hojně používanou metodou, která pomocí konvoluce sledu akčních potenciálů gaussovským jádrem odhaduje hledanou funkci [3], [4], [5], [6], [7], [8]. Nejproblematictější částí této metody je výběr šířky jádra, protože podstatně ovlivňuje výslednou kvalitu odhadu. Častým řešením je pak manuální odhad šířky, protože není jasné vodítko, která ji stanovit exaktně. V nedávné práci [1] autoři navrhli algoritmus, jak za určitých podmínek přesně vypočítat optimální šířku.

Pro větší množiny vstupních dat roste časová složitost přímočaré verze algoritmu natolik, že je nemožné ji používat v interaktivním režimu pro studium metody nebo dat samých. V tomto příspěvku je navrženo řešení, které snižuje časovou složitost na únosnou míru i pro rozsáhlejší množinu vstupních dat. Samotné vylepšení nijak nemění výsledné řešení původní metody – vlastnosti řešení a srovnání s jinými metodami lze najít v původním článku [1].

2. Metody

2.1 Původní metoda

Firing rate je nezáporná funkce λ , pro kterou se integrál $\int_b^a \lambda(t) dt$ rovná střední hodnotě pozorovaných akčních potenciálů během časového intervalu $[a, b]$. V klasickém experimentu dostaneme jeden nebo více záznamů sledu akčních potenciálů a úlohou je odhadnout funkci λ' , co možná nejlíže k původní λ (tj. funkci, o které věříme, že stojí v pozadí generování akčních potenciálů). V studované metodě se použije konvoluce sledu akčních potenciálů vybraným jádrem pro získání vyhlazené křivky, která aproximuje λ . V našem případě používáme gaussovské jádro

$$k_w(t) = \frac{1}{\sqrt{2\pi}w} \exp\left(-\frac{t^2}{2w^2}\right)$$

a problém se redukuje na otázku, jak vybrat optimální šířku w , aby rozdíl mezi λ a λ' byl co nejmenší. Metoda samotná přesahuje možnosti tohoto článku, detaily je možné najít v [1]. Pro naše potřeby je důležité, že jádro výpočtu lze shrnout následující úlohou: najdi hodnotu w^* tak, aby funkce (1) nabývala minimální hodnotu:

$$C(w) = \frac{1}{n^2} \sum_{i,j} \psi_w(t_i, t_j) - \frac{2}{n^2} \sum_{i \neq j} k_w(t_i - t_j) \quad (1),$$

Kde t_i je čas i -tého akčního potenciálu, n je počet záznamů,

$$\psi_w(t_i, t_j) = \int_b^a k_w(t - t_i) k_w(t - t_j) dt$$

$[a, b]$ definuje časový interval záznamu a k_w je použité jádro. Protože jsme zde vybrali Gaussovské jádro, lze (1) přepsat jako

$$2\sqrt{\pi} n^2 C(w) = \frac{N}{w} + \frac{2}{w} \sum_{i < j} \left(e^{-\frac{(t_i - t_j)^2}{4w^2}} - 2\sqrt{2} e^{-\frac{(t_i - t_j)^2}{2w^2}} \right) \quad (2),$$

kde N je počet akčních potenciálů. Povšimněme si rovněž, že člen $2\sqrt{\pi} n^2$ je konstantní a nemá na výsledné w^* vliv. Pomocí $E(w, t_i, t_j)$ označme vnitřní členy sumy v (2). Dále označme \mathcal{W} prostor možných hodnot w , ve kterém vyhledáváme w^* . Dále označme jeho velikost $W = |\mathcal{W}|$ a předpokládejme, že jednotlivé hodnoty mají od sebe stejnou vzdálenost. Přímochará implementace bude hledat minimum pomocí tří vnořených smyček:

(11) pro $w \in \mathcal{W}$

(12) pro $j \in [1, \dots, N]$

(13) pro $i \in [1, \dots, j-1]$

$Tmp = E(w, t_i, t_j)$

if ($min > tmp$) $min = tmp$, $w^* = w$

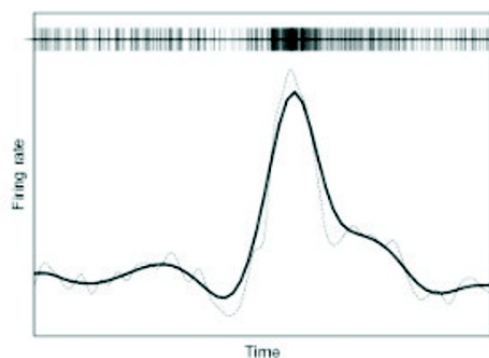
což dává časovou složitost $O(N^2W)$. Volba W je závislá na intervalu $[a, b]$ a požadované přesnosti optimální hodnoty. V typickém případě $w^* \ll b - a$ a je možné pro horní mez W volit $\log(b - a)$, v případě dělení intervalů (viz níže) horní mez nehraje zásadní roli.

2.2 Dělení intervalů

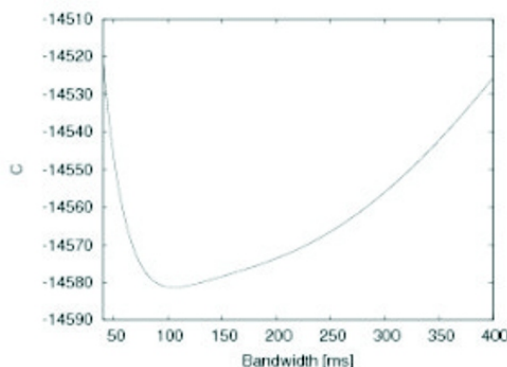
V typickém případě je C unimodální funkcí, viz Obr. 2. (toto tvrzení však nelze zobecnit – problém by kupříkladu nastal v případě hledání optima s rozlišením nižším než je vzorkovací frekvence měřených dat). Máme-li rozumné hodnoty pro horní a dolní mez, pak je možné pro unimodální funkce použít některý z algoritmů pro hledání minima, založených na rekurzivním dělení intervalů. To zredukuje časovou složitost smyčky (I1) na logaritmický faktor, úhrnem pak dostáváme složitost $O(N^2 \log(W))$. Jak naznačeno výše, určitá pozornost musí být věnována průběhu funkce, před přímočarým použitím dělení.

Vzhledem k tomu, že $E(w, t_i, t_j)$ je člen nezávislý na předchozích výpočtech, je poměrně přirozené úlohu paralelizovat pro běh na více výpočetních jednotkách. Rozložení úlohy do paralelních bloků na úrovni smyčky (I1) by nám nedovolilo použít paralelizaci v případě metody dělení intervalů, proto zvolíme dělení na úloh v úrovni (I2). Tím dostáváme celkový odhad časové složitosti $O(N^2 \log(W)/p)$.

Nyní přesuňme naši pozornost k implementaci.



Obr. 1. Ilustrační příklad. Z prvotní funkce λ (tlustou čarou) je generován experimentálně měřený sled akčních potenciálů (horní linka), ze které se zpětně snažíme odvodit optimální λ' (tenkou čarou).

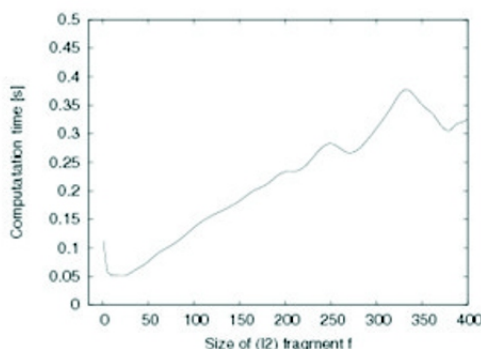


Obr. 2. Typický tvar funkce C získané pro experimentální data.

2.3.1 Rozklad úloh

Protože horní mez ve smyčce (I3) není konstantní, triviální rozklad úrovně (I2) nagenereuje p podúloh $[1, \dots, N/p], [N/p, \dots, 2N/p], \dots, [(p-1)N/p, \dots, N]$ s narůstající časovou složitostí pro každou z podúloh. Nakonec budou výpočetní jednotky prvních podúloh nečinně čekat na dokončení výpočtu posledních.

To lze řešit vícero způsoby – (1) přesunutím rozkladu podúloh do úrovně (I3), (2) rozkladem p podúloh v úrovni (I2) proporcionálním způsobem, aby každá podúloha zabírala stejný procesorový čas anebo (3) rozklad úrovně (I2) do řádově více malých podúloh, které si postupně rozebírá p jednotek. Ve skutečné implementaci jsme vybrali (3), jelikož (1) má tendenci produkovat přebytkovou zátěž v samotném distribučním mechanismu podúloh a (2) předpokládá, že výpočetní jednotky jsou ekvivalentní co do výkonu i dosažitelnosti (to v mnoha případech nemusí platit).



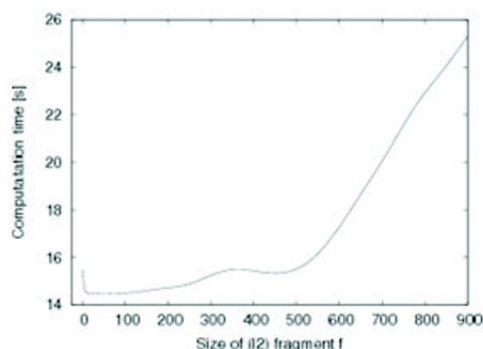
3. Výsledky

3.1 Ladění parametrů

Na základě předchozích odstavců byl implementován algoritmus podporující popsané strategie – přímočarý průchod i dělení, obě v jak v sekvenční, tak v paralelní verzi. Zvoleným jazykem bylo C++, pro paralelizaci bylo použita openMPI implementace [3] MPI standardu, pro dělení byl použit Brentův minimalizační algoritmus [4]. Abychom našli optimální rozklad podúloh, provedli jsme analýzu časových nároků pro rozdílnou fragmentaci podúloh, viz Obr. 3.

Analýza proběhla pro dvě vstupní množiny dat – 1000 a 18000 akčních potenciálů. V případě větší z množin vidíme, že libovolná hodnota pod $f=500$ má přibližně stejnou časovou složitost. Vrcholek na samotném začátku je způsoben růstem nároků paralelizačních služeb (tj. cena distribuce podúloh začne spotřebovávat více zdrojů nežli výpočet podúloh samotný). Zhruba od $f=500$ lze vidět postupný nárůst času potřebného pro celý výpočet, jenž je způsoben nedostatečnou fragmentací (tj. některé jednotky zbytečně čekají na dokončení výpočtu ostatních jednotek). Snižováním počtu AP na vstupu bude tato hodnota klesat, jak též vidno na výsledcích pro 1000 AP. Tvar křivky se zachovává, ale celkové časové nároky na výpočet jsou zanedbatelné.

To vedlo k výběru optimální velikosti fragmentu $f=100$, který bude dostatečný pro jakékoliv větší vstupy. Jak lze vidět



Obr. 3. Obrázek ilustruje, jakým způsobem se projevuje fragmentace paralelního zpracování na celkovém času výpočtu. Vstupní data pro levý panel činila 1000 AP, pro pravý 18000 AP. Oba vstupy pocházely z experimentálně naměřených dat, pro výpočet byl zvolen algoritmus dělení. f značí velikost jedné podúlohy, tj. počet iterací smyčky (I3).

Tab. 1. Porovnání časových nároků.

Metoda	Čas (min:sec)
Sekvenční prohledávání	58:41
Paralelní prohledávání	4:14
Sekvenční dělení	2:43
Paralelní dělení	0:09

z levého panelu na Obr. 3, je to rozumná volba i pro malou velikost vstupu, byť ne příliš důležitá vzhledem k celkovému času výpočtu. Samozřejmě je tato hodnota závislá na konkrétní konfiguraci výpočetního prostředí – v našem případě se všechny testy odehrávaly na clusteru s dvaceti procesorovými jádry.

3.2 Konkrétní časové nároky

Pro srovnání času potřebného na výpočet uvádíme Tabulku 3. Vstupní data a parametry byly shodné pro všechny úlohy: 18000 AP, interval pro hledanou šířku [1,400] ms, přesnost 1 ms ($W=400/1$).

4. Diskuze

Navrhli jsme a implementovali paralelní algoritmus pro hledání optimální šířky kernelu, který zlepšuje časovou složitost oproti jeho přímočaré verzi. Pokud je navíc ohodnocující funkce (1) unimodální na

daném intervalu, lze použít algoritmus dělení, který snižuje časovou složitost ještě dramatičtěji. Odhadnout rozumné meze pro interval lze předběžným průchodem, který používá malou vzorkovací množinu.

Toto zlepšení nehraje žádnou roli v případě malé vstupní množiny dat, je však výrazné pro rozsáhlejší vstup. Řešení samotného problému bylo motivováno prací se vstupy o velikosti řádově ~20000 akčních potenciálů, kdy bylo zároveň zapotřebí vyhodnocovat i odlišné podmnožiny vstupu – i přibližná znalost ohodnocující funkce dovolila použití dělicí verze algoritmu.

Na závěr jsme stanovili parametry pro běh algoritmu v prostředí konkrétního výpočetního clusteru a změřili přesné časové nároky jednotlivých strategií.

Poděkování

Tento výstup vznikl v rámci projektu Specifického vysokoškolského výzkumu SVV-2010-265 513.

Literatura

- [1] Shimazaki H., Shinomoto S.: Kernel bandwidth optimization in spike rate estimation. *Journal of Computational Neuroscience* 2009; 29: 171182.
- [2] Cunningham J.P., Gilja V., Ryu S.I., Shenoy, K.V.: Methods for estimating neural firing rates, and their application to brain-machine interfaces. *Neural Networks* 2009; 22 (9): 12351246.
- [3] Sanderson A.C.: Adaptive Filtering of Neuronal Spike Train Data. *IEEE*

Transactions on Biomedical Engineering 1980; 27 (5): 271 274.

- [4] Richmond B.J., Optican L.M., Podell M., Spitzer H.: Temporal encoding of two-dimensional patterns by single units in primate inferior temporal cortex. I. Response characteristics. *J Neurophysiol* 1987; 57 (1): 132146.
- [5] Richmond B.J., Optican L.M., Spitzer H.: Temporal encoding of twodimensional patterns by single units in primate primary visual cortex. I. Stimulus-response relations. *J Neurophysiol* 1990; 64 (2): 351369.
- [6] Paulin M.G.: Digital filters for firing rate estimation. *Biological cybernetics* 1992; 66 (6): 525531.
- [7] Paulin M.G., Hoffman L.F.: Optimal firing rate estimation. *Neural Networks* 2001; 14 (6-7): 877 881.
- [8] Nawrot M., Aertsen A., Rotter S.: Single-trial estimation of neuronal firing rates: From single-neuron spike trains to population activity. *Journal of Neuroscience Methods* 1999; 94 (1): 81 92.
- [9] Gabriel E., Fagg G.E., Bosilca G., Angskun T., Dongarra J.J., Squyres J.M. et al.: Open MPI: Goals, Concept, and Design of a Next Generation MPI Implementation. In: *Proceedings, 11th European PVM/MPI Users' Group Meeting* 2004; 97104.
- [10] Brent R.P.: *Algorithms for minimization without derivatives*. Dover Pubns; 2002.

Kontakt

Mgr. Pavel Šanda

Fyziologický ústav AV ČR, v.v.i.
Vídeňská 1083
142 20 Praha 4
e-mail: ps@ucw.cz

Jednoduchá ekonomická analýza diagnostiky nemocí štítné žlázy při screeningu těhotných žen

Z. Telička¹, J. Jiskra¹, D. Springer², Z. Límanová¹

¹3. interní klinika 1. lékařské fakulty a Všeobecné fakultní nemocnice v Praze

²Ústav klinické biochemie a laboratorní diagnostiky 1. lékařské fakulty a Všeobecné fakultní nemocnice v Praze

Školitel: Prof. MUDr. Štěpán Svačina, DrSc., MBA

Souhrn

Onemocnění štítné žlázy jsou u těhotných žen časté. V České republice není zaveden plošný screening na toto onemocnění. Cílem práce bylo zhodnotit náklady pilotní studie screeningu onemocnění štítné žlázy u těhotných žen.

Použili jsme jednoduché statistické metody pro srovnání nákladů na pilotní screeningovou studii 2651 žen (vyšetřeny parametry TSH a TPOAb) vzhledem ke skutečným současným nákladům dle číselníku výkonů. Provedli jsme analýzu nákladů a přínosů pro pacientky v případě, že se vyšetřuje pouze TSH nebo TPOAb parametr.

Screening pouze TSH parametru je 2,7 krát nákladově efektivnější než TPOAb. Screening pouze TSH může vést k vynechání 77 žen pozitivně testovaných na TPOAb. Screening pouze TPOAb parametru může znamenat vynechání 77 žen pozitivně testovaných na TSH.

K dosažení lepší senzitivity je nutné provádět TSH a TPOAb screening společně, protože oba parametry se vzájemně vhodně doplňují, ačkoliv TSH screening je 2,7 krát levnější.

Klíčová slova: onemocnění štítné žlázy, tyreopatie, těhotenství, nákladová efektivita, screening, iq, qaly

1. Úvod

Onemocnění štítné žlázy jsou u těhotných žen častá [1]. Neléčená dysfunkce štítné žlázy v těhotenství má negativní vliv jak na matku, tak i na dítě. Může způsobit nízké koncentrace hormonů štítné žlázy, které jsou důležité pro správný vývoj mozku dítěte, zvláště v začátku těhotenství [2]. Screeningové testy onemocnění štítné žlázy se skládají z testování hladin protilátek proti peroxidáze štítné žlázy,

tyreoidu stimulující hormonu (TSH) a hladin volného tyreoidálních hormonů (T3 a T4) v krvi. Zvýšená hladina TPOAb může indikovat autoimunitní reakci, při které tělo posílá imunitní buňky proti štítné žláze. V zahraničních studiích má asi 11 % těhotných žen prokázáno onemocnění štítné žlázy po vyšetření krve [1]. Observační studie zjistily, že léčba mateřské hypotyreoidy tyreoidálním hormonem snižuje incidenci gestační hypertenze u matek [3] a zlepšuje IQ u dětí [4], [5]. Pozitivní TSH výsledky (zároveň negativní TPOAb výsledky) mohou vést k nesprávnému vývoji mozku dítěte v průběhu těhotenství a pozitivní hodnoty TPOAb (a zároveň negativní výsledky TSH) mohou vést k poporodnímu onemocnění štítné žlázy. Vysoké hodnoty incidence vedou lékaře k úvahám zavedení systematického screeningu na toto onemocnění těhotných žen. V současné době probíhá diskuze, zda tento screening bude vhodné zavést celoplošně nebo jen pro určité rizikové skupiny, a které parametry vyšetřovat. Všechny tyto faktory jsou také postihnuty ekonomickými aspekty ze dvou stran:

- 1) Budou náklady na screening adekvátní k počtu pozitivních nálezů?
- 2) Jaký to bude mít přínos pro pozitivně vyšetřenou matku a dítě, která pak podstoupí léčbu tohoto onemocnění?

1.1 Již proběhla zahraniční studie

Analýza nákladové efektivity na screening onemocnění štítné žlázy u těhotných žen již proběhla v USA. Byly použity tři strategie, které byly aplikovány na testované skupiny: ženy bez screeningu, jednorázový screening na TSH a jednorázový screening na TPOAb. Abnormální výsledky screeningových testů byly následovány dalšími testy a hypotyreózní ženy podstoupily léčbu levothyroxinem, dokud se výsledky nevrátily do normálních

hodnot. Nakonec byla provedena nákladová efektivita, která zahrnovala náklady na:

- návštěvy v nemocnici a u endokrinologa,
- laboratorní testy na TSH nebo TPOAb,
- léčba levothyroxinem.

V této studii byly zjištěny náklady na nižší IQ, které se projevilo u dětí. Autoři projektu pracovali s faktem, že je signifikantní pokles IQ dítěte, pokud se matky neléčily z onemocnění štítné žlázy již v průběhu raného stádia těhotenství. Rozdíl v jednom bodu IQ byl asociován s poklesem příjmu okolo 2 % [7]. S touto hypotézou vypočítali pokles v budoucích příjmech dítěte v hodnotě 15 513 USD na jeden IQ bod. Na základě Hadowovy follow-up studie [8] zjistili, že děti neléčených matek se subklinickou hypotyreózou mají průměrně pokles 9,2- až 4,2 bodů IQ (po upravení 0,8 body v socioekonomickém statusu ve srovnání s kontrolami). Vytvořením Markovova Modelu [9] bylo možné zjistit hodnotu inkrementální nákladové efektivity u dítěte na jeden rok kvalitního života QALY (quality adjusted life year), pokud je žena léčena v průběhu těhotenství ve srovnání s těmi neléčenými [6]. Screening těhotných žen v prvním trimestru na TSH bylo nákladově efektivnější v porovnání se ženami bez screeningu, zatímco screening na ATPOAb v porovnání s TSH screeningem byla nákladově efektivnější v ceně 15 182 USD / QALY.

1.2 První studie provedená v České republice

V České republice začala v roce 2009 multicentrická pilotní studiem ve 13 regionech země, zaměřila se na screeningové testy onemocnění štítné žlázy u těhotných žen a jejím cílem bylo zjistit první data, která by mohla lékařům a pojišťovnám

pomoci zjistit, jestli je tento screening pro matku a dítě přínosný a také nákladově snesitelný pro systém zdravotního pojištění. V naší studii jsme se pokusili vyčíslit náklady na tuto studii, ve které byly testovány TSH a TPOAb parametry. Také jsme se pokusili modelovat náklady na screening, pokud by byl proveden pouze s TSH nebo TPOAb parametrem samostatně.

2. Metody

V průběhu 11.-13. týdne těhotenství byly testovány ženy na TSH a TPOAb. Pokud byly laboratorní výsledky pozitivní, ženám bylo doporučeno navštívit endokrinologa, který eventuálně zahájil léčbu. Náklady na provedení laboratorních výsledků jsou 518 Kč (TSH 140 Kč; TPOAb 378 Kč). Vzhledem k tomu, že v současné době neexistují standardizované screeningové testy na TSH a TPOAb, každá laboratoř používá vlastní testovací metodu s rozdílnými mezemi. Orientační normální meze pro TSH jsou 0,1-3,7 mIU/l a pro TPOAb ≤34 IU/ml. Ze studie byly vyjmuty všechny pacientky pozitivně testované na TSH kvůli supresi TSH, abychom vyloučili zdravé ženy s fyziologickou hladinou TSH v průběhu prvního trimestru těhotenství. Analyzovali jsme proto pacientky se zvýšenou hodnotou TSH (>3,7 mIU/l). Vypočítali jsme celkové finanční náklady na provedení pilotního screeningu a podle počtu pozitivních výsledků u jednotlivých parametrů jsme následně vypočítali náklady na zachytí jedné pozitivní diagnózy.

3. Výsledky

Náklady na screening vyšetření TSH a zároveň TPOAb parametru byly 1 373 218 Kč. 612 (23 %) pacientek bylo vyhodnoceno jako pozitivní v alespoň jednom z testovaných parametrů. To znamená náklady na provedení jednoho pozitivního testu 2 243 Kč.

Také jsme chtěli zjistit, jestli by bylo dostačující testovat ženy pouze na TSH nebo na TPOAb parametr. Pozitivně testovaných (nad referenční interval) na TSH bylo zjištěno 269 žen (10,14 %) a z toho 77 (28,63 %) bylo pozitivní na TPOAb a 192 (71,37 %) bylo TPOAb negativní. Náklady na tento screening byly 371 140 Kč, tzn. 1 380 Kč na jeden pozitivní TSH test. Ale při 2 382 (89,85 %)

negativně testovaných žen na TSH bylo z toho 143 (6 %) TPOAb pozitivní. To znamená, že pokud by byl proveden pouze TSH screening, mohlo by to vést k nezachycení 77 žen s TPOAb pozitivními výsledky.

Pokud se provedl pouze screening na TPOAb, pozitivních záchytů bylo 220 (8,3 %), z toho 77 (35 %) žen mělo pozitivní TSH výsledky. Náklady na tento TPOAb screening by byly 1 002 078 Kč, tzn. 4 555 Kč na jeden TPOAb pozitivní screening. Ale 2 431 (91,70 %) TPOAb negativních žen z toho mělo 192 (7,9 %) TSH pozitivní screeningový test. To znamená, že screening pouze TPOAb parametru může vést k nezachycení 77 pacientek s TSH pozitivními výsledky.

4. Diskuze

Z výsledků jsme zjistili, že oba parametry, jak TSH, tak TPOAb jsou během screeningu nezastupitelné a měly by být vyšetřovány zároveň, přestože TSH screening je 2,7 krát levnější než TPOAb screening. Přitom screening pouze jednoho z uvedených parametrů by mohl znamenat vynechání 77 žen, které mohou být také pozitivní.

Navíc, pozitivní screening TSH nebo TPOAb mohou znamenat pro matku a dítě rozdílná rizika: Zatím co pozitivní TSH výsledky mohou vést k nesprávnému vývoji mozku dítěte v průběhu těhotenství, tak pozitivní hodnoty TPOAb mohou vést k poporodnímu onemocnění štítné žlázy.

Poděkování

Tato práce byla podpořena grantem Ministerstva zdravotnictví IGA NS10595-3 a projektem SVV-2010-265513.

Literatura

- [1] Hollowell J.G., Staehling N.W., Flanders W.D., Hannon W.H., Gunter E.W., Spencer C.A., Braverman L.E.: Serum TSH, T(4), and thyroid antibodies in the United States population (1988 to 1994): National Health and Nutrition Examination Survey (NHANES III). *J Clin Endocrinol Metab.* 2002 Feb;87(2):489-99.
- [2] Springer D., Zima T., Limanova Z.: Reference intervals in evaluation of maternal thyroid function during the first

trimester of pregnancy. *Eur J Endocrinol.* 2009 May;160(5):791-7. Epub 2009 Feb 19.

- [3] Leung A.S., Millar L.K., Koonings P.P., Montoro M., Mestman J.H.: Perinatal outcome in hypothyroid pregnancies. *Obstet Gynecol.* 1993 Mar;81(3):349-53.
- [4] Haddow J.E., Palomaki G.E., Allan W.C., Williams J.R., Knight G.J., Gagnon J., O'Heir C.E., Mitchell M.L., Hermos R.J., Waisbren S.E., Faix J.D., Klein R.Z.: Maternal thyroid deficiency during pregnancy and subsequent neuropsychological development of the child. *N Engl J Med.* 1999 Aug 19;341(8):549-55.
- [5] Man E.B., Brown J.F., Serunian S.A.: Maternal hypothyroxinemia: psychoneurological deficits of progeny. *Ann Clin Lab Sci.* 1991 Jul-Aug;21(4):227-39.
- [6] Dosiou C., Sanders G.D., Araki S.S., Crapo L.M.: Screening pregnant women for autoimmune thyroid disease: a cost-effectiveness analysis. *Eur J Endocrinol.* 2008 Jun;158(6):841-51.
- [7] Grosse S.D., Matte T.D., Schwartz J., Jackson R.J.: Economic gains resulting from the reduction in children's exposure to lead in the United States. *Environ Health Perspect.* 2002 Jun;110(6):563-9.
- [8] Klein R.Z., Sargent J.D., Larsen P.R., Waisbren S.E., Haddow J.E., Mitchell M.L.: Relation of severity of maternal hypothyroidism to cognitive development of offspring. *J Med Screen.* 2001;8(1):18-20.
- [9] Sonnenberg F.A., Beck J.R.: Markov models in medical decision making: a practical guide. *Med Decis Making.* 1993 Oct-Dec;13(4):322-38.

Kontakt

Mgr. Zdeněk Telička

3. interní klinika 1. lékařské fakulty a
Všeobecné fakultní nemocnice v Praze
U Nemocnice 1
128 08 Praha 2
e-mail: zdenek@telicka.cz

Identifikace jednotek a dalších termínů v českých lékařských zprávách

K. Zvára, V. Kašpar¹

¹Centrum biomedicínské informatiky, Ústav informatiky AV ČR, v.v.i.

Školitel: Doc. Ing. Vojtěch Svátek, Dr.

Souhrn

Zdravotnická dokumentace má v České republice obvykle povahu volného textu formátovaného jen pomocí mezer, tabulátorů a nových řádků. Extrakce informací z takových zpráv představuje výzvu, jejíž splnění by umožnilo její snadné využívání při poskytování zdravotní péče v zahraničí či převod do strukturované formy. Pro identifikaci a extrakci informací lze využít různých přístupů, od regulární analýzy po pokročilou lingvistickou analýzu a statistické metody. Článek shrnuje naše poznatky získané při regulární analýze a při využití běžně využívaných klasifikačních systémů.

Klíčová slova: zpracování přirozeného jazyka, zdravotnická dokumentace, elektronický zdravotní záznam, regulární analýza, lingvistická analýza

1. Zdravotnická dokumentace a lékařská zpráva v ČR

Náležitosti zdravotnické dokumentace, a dokonce ani povinnost ji vést, nebyly v českém zdravotnictví dlouho upraveny. Právní otázky v souvislosti se zdravotnickou dokumentací upřesňuje například JUDr. MUDr. Roman Žďárek, Ph.D. v [1]. V roce 2001 však byl přijat zákon č. 260/2001 Sb., který novelizoval zákon č. 20/1966 Sb., o péči o zdraví lidu, a vložil do něj nové paragrafy 67a až 67d upravující otázky zdravotnické dokumentace a národního zdravotnického informačního systému. Zákon výslovně stanovil, že vedení zdravotnické dokumentace, jakož i další nakládání s ní, je zpracováním osobních údajů ve smyslu zákona č. 101/2000 Sb., o ochraně osobních údajů.

Podle § 67b, odst. 2 zákona o péči o zdraví lidu, v platném znění, zdravotnická dokumentace obsahuje:

- osobní údaje pacienta v rozsahu nezbytném pro identifikaci pacienta a zjištění anamnézy,
- informace o onemocnění, o průběhu a výsledcích vyšetření, léčení a dalších významných okolnostech souvisejících se zdravotním stavem pacienta a s postupem při poskytování zdravotní péče.

Podle ustanovení § 67b odst. 4 zákona o péči o zdraví lidu, v platném znění, zápis ve zdravotnické dokumentaci musí být veden průkazně, pravdivě a čitelně; je průběžně doplňován a musí být opatřen datem zápisu, identifikací a podpisem osoby, která zápis provedla. Opravy ve zdravotnické dokumentaci se provádí novým zápisem s uvedením dne opravy, identifikací a podpisem osoby, která opravu provedla. Původní záznam musí zůstat čitelný.

V Příloze 1 k vyhlášce 385/2006 Sb. je uveden požadavek na minimální rozsah části zdravotnické dokumentace, která poskytuje informace o provedeném vyšetření. Tato část zdravotnické dokumentace se nazývá lékařská zpráva a musí obsahovat alespoň následující informace: a) údaje o zjištěném zdravotním stavu, včetně výsledků laboratorních a dalších vyšetření, b) údaje o dosavadní léčbě a reakci pacienta na ní, c) doporučení k dalšímu poskytování zdravotní péče.

České lékařské zprávy mají obvykle formu volného textu. Skutečnost, že lékaři jsou organizováni v profesní organizaci a mají obdobné vzdělání, vede k tomu, že zprávy psané různými lékaři mají obdobnou strukturu, především pořadí jednotlivých částí.

Text je obvykle formátován pouze pomocí mezer a nových řádků, žádné značkování se obvykle nepoužívá. Lékařské zprávy vznikají buď od začátku (napsáním celé zprávy lékařem), ze šablony (poskytující automaticky generované údaje jako jsou identifikace pacienta, jeho zdravotní pojišťovna, jméno lékaře a dnešní datum) či úpravou starší lékařské zprávy stejného pacienta. Vzhledem k ušetřené námaze při vypisování administrativních údajů (v případě tvorby ze šablony) a při zapisování údajů z osobní a rodinné anamnézy (při tvorbě zprávy úpravou starší zprávy téhož pacienta), jsou tyto dva způsoby častější než vytváření zprávy od úplného začátku (z prázdného dokumentu). Využívání starších lékařských zpráv a jejich editace popisující aktuální zdravotní stav pacienta však může vést k závažným opomenutím, kdy v nové zprávě zůstávají omylem i části předchozí zprávy, které však již do popisu aktuálně provedeného vyšetření nepatří a v některých případech jsou zavádějící. Tyto problémy s kopírováním a aktualizací starších lékařských zpráv jsou pozorovány i v jiných zemích a podrobně je popisuje například práce [2].

Dalším specifickým českých lékařských zpráv je občasné vkládání textového obsahu z cizích zdrojů. Takovým obsahem bývá např. výsledek biochemického vyšetření, který může být vložen ve formě textově formátované tabulky, přičemž formátování se zachová v případě, že je použito neproporcionální písmo. Údaje z biochemického vyšetření tak jsou uchovány zpravidla v řádcích, kde každá buňka představuje hodnotu parametru, například v pořadí: „název testu“, „spodní referenční mez“, „horní referenční mez“, „zjištěná hodnota“ a „grafické znázornění vztahu zjištěné hodnoty k referenčním mezím“.

České lékařské zprávy obsahují velké množství překlepů, zkratk a zkrácených slov viz například [3]. Jednotlivé použité zkratky i zkrácená slova přitom dávají různý smysl v různých kontextech. V lékařských zprávách často chybí interpunkce či mezery. Odborné termíny různí lékaři označují různě, například „diabetes mellitus“ je někdy zkracován jako „diab.mel.“, jindy jako DM. Údaje anamnézy přitom nejsou uvedeny jako prostý výčet diagnóz, ale často ve větě, kde větná stavba samotná nese další informaci. Stejně tak část popisující předepsanou terapii může nést informativní obsah určený přímo pacientovi; např. „musí-li mít pac. denně polévku a ještě hlavní jídlo, pak do polévky vždy otruby – zde v Biocentru“.

Kromě termínu zdravotnická dokumentace se často setkáváme i s termínem zdravotní dokumentace. Termín zdravotní dokumentace jsme nenalezli v žádném oficiálním dokumentu vydaném MZ ČR. V praxi se s ním ale setkáváme poměrně často, zejména na různých webových stránkách. Z našeho pohledu termín zdravotní dokumentace pokrývá jenom část informací ze zdravotnické dokumentace, konkrétně část obsahující informace o zdravotním stavu pacienta. Tato interpretace zdravotní dokumentace by odpovídala nástrojům pro sběr informací o zdravotním stavu pacienta pomocí elektro-nického zdravotního záznamu.

2. Účely identifikace terminů a jednotek

Plnotextová zdravotnická dokumentace v dnešní podobě dostává potřebám lékařů hovořících stejným jazykem pro poskytování zdravotní péče, přičemž využití údajů ze zdravotnické dokumentace je omezeno na jejich interpretaci, případně na ruční vkládání do podpůrných systémů.

Lékaři hovoří jiným jazykem než lékař, který zdravotnickou dokumentaci pořídil, však nemohou samostatně současnou textovou zdravotnickou dokumentaci využít. Pro její využití potřebují odborníka, který ovládá jak jazyk zdravotnické dokumentace pacienta, tak jazyk, kterému rozumí současný ošetřující lékař. Lékaři tak často využívají jen údaje získané od pacienta a údaje ze zdravotnické doku-

mentace, o kterých se domnívají, že jim rozumí. Porozumění v takovém případě napomáhá společný (latinský) základ lékařských názvů v západním světě a očekávaný formát a rozsahy hodnot (například u krevního tlaku). Jedním z praktických použití metod pro extrakci informací by tedy mohl být jejich přepis do jiného jazyka.

Odborné a standardizační organizace zabývající se medicínskou informatikou již dlouhou dobu připravují standardy pro strukturovaný zápis zdravotních údajů o pacientovi. Mezi takové standardy patří i HL7 CDA a EN 13606 (EHRcom). Dalším praktickým použitím extrahovaných informací tedy může být převádění textové zdravotnické dokumentace do strukturované formy, ať již značkováním textu (jako u HL7 CDA) či do samostatné grafové struktury (jako v EN 13606).

Údaje ze zdravotnické dokumentace je výhodné využívat i pro další účely. Může jít o sloučení zaznamenávání poskytnuté zdravotní péče pro účely poskytování další péče se zaznamenáváním za účelem jejího vyúčtování (pacientovi, zdravotním pojišťovnám apod). Stejně tak mohou s výhodami údaje ze zdravotnické dokumentace využívat podpůrné systémy jako automaty sledující, zda je poskytovaná péče slučitelná s určenými postupy (např. s lékařskými doporučeními). Vidíme, že data ze zdravotnické dokumentace lze využít k řadě dalších účelů. Velmi podrobně možnosti dalšího využití dat ze zdravotnické dokumentace analyzuje práce [4].

Všech tří hlavních, výše uvedených účelů identifikace údajů ve zdravotnické dokumentaci ve formě volného textu lze dosáhnout prostřednictvím extrakce informací do strukturovaného zdravotního záznamu s grafovou strukturou. Velmi důležitá je zejména extrakce těch informací, které jsou velmi užitečné při poskytování neodkladné péče.

Takovými informacemi jsou především:

- seznam známých alergií, tento bývá součástí osobní anamnézy,
- seznam současných a dřívějších diagnóz včetně poslední hodnoty krevního tlaku a výsledků biochemických vyšetření,

- identifikační údaje, především jméno, příjmení a číslo pojištěnce (rodné číslo).

Pro tyto účely bohatě postačuje struktura EN 13606, Mezinárodní klasifikace nemocí (MKN10) a vybrané kódy z číselníku LOINC příp. NČLP (Soubor číselníků NČLP (Národní číselník laboratorních položek), <http://ciselniky.dasta.mzcr.cz>).

3. Extrakce informace z lékařských zpráv

Extrakci informace z lékařských zpráv pomocí nestatistických metod se věnoval Jiří Semecký [5]. Uvedl dva přístupy k sémantické analýze volného lékařského textu. V prvním přístupu jde o nalezení algoritmu pro vyhledání fragmentů lékařské zprávy, které mají význam popsáný znalostní bází. V druhé definici jde o nalezení algoritmu pro sémantickou analýzu a označení nalezených fragmentů ve zprávě značkami. V práci [5] jsou uvedeny důvody, pro které se zdá, že lingvistická analýza lékařských zpráv nemůže být úspěšná. Mezi tyto důvody patří skutečnost, že lékařské zprávy obvykle neobsahují věty, nejsou správně odděleny interpunkcí a tvrzení mají povahu hesel. Jako druhý způsob analýzy volného lékařského textu práce [5] uvádí regulární analýzu, tedy ověřování, zda jednotlivé úseky zprávy splňují předem určená, často na jazyku nezávislá pravidla.

Peter Smatana [6] navazuje na práci [5] a pokračuje v experimentech s lingvistickou a sémantickou analýzou. Obdobně jako v práci [5] ani v práci [6] nejsou použity žádné databáze možných terminů ani obecné jazykové slovníky. Při rozšíření o použití lingvistické analýzy dochází k mírně lepším výsledkům než při využití jen analýzy pomocí regulárních výrazů. Přesto, že práce [2] je psána ve slovenském jazyce, jejím podkladem byly zprávy psané česky.

Velmi palčivým problémem při rozpoznávání významu obsahu v lékařských zprávách je časté zkracování slov. Problematice se v rámci aplikace metod pro zpracování přirozeného jazyka věnovali různí autoři [7], [8].

Pro rozpoznávání zkratk navrhovali vytváření zvláštních slovníků za možné asistence lingvistické analýzy. Zajímavé tvrzení o zkratkách v medicíně přinesl Tsong O. Cheng [7] v části "Words to the Editor" časopisu Journal of the Royal Society of Medicine. V příspěvku uvádí, že více než 90 % začínajících lékařů z jiných specializací nerozumí zkratkám použitých v lékařských zprávách. Přestože jeho příspěvek přímo nesouvisí s problematikou zpracování přirozeného jazyka, dokládá tímto tvrzením velkou nejednoznačnost a značnou kontextovou závislost významu zkratk.

V následující části uvedeme náš přístup k identifikaci číselných hodnot a terminů v lékařských zprávách. K ověření navržených postupů bylo použito celkem 107 lékařských zpráv.

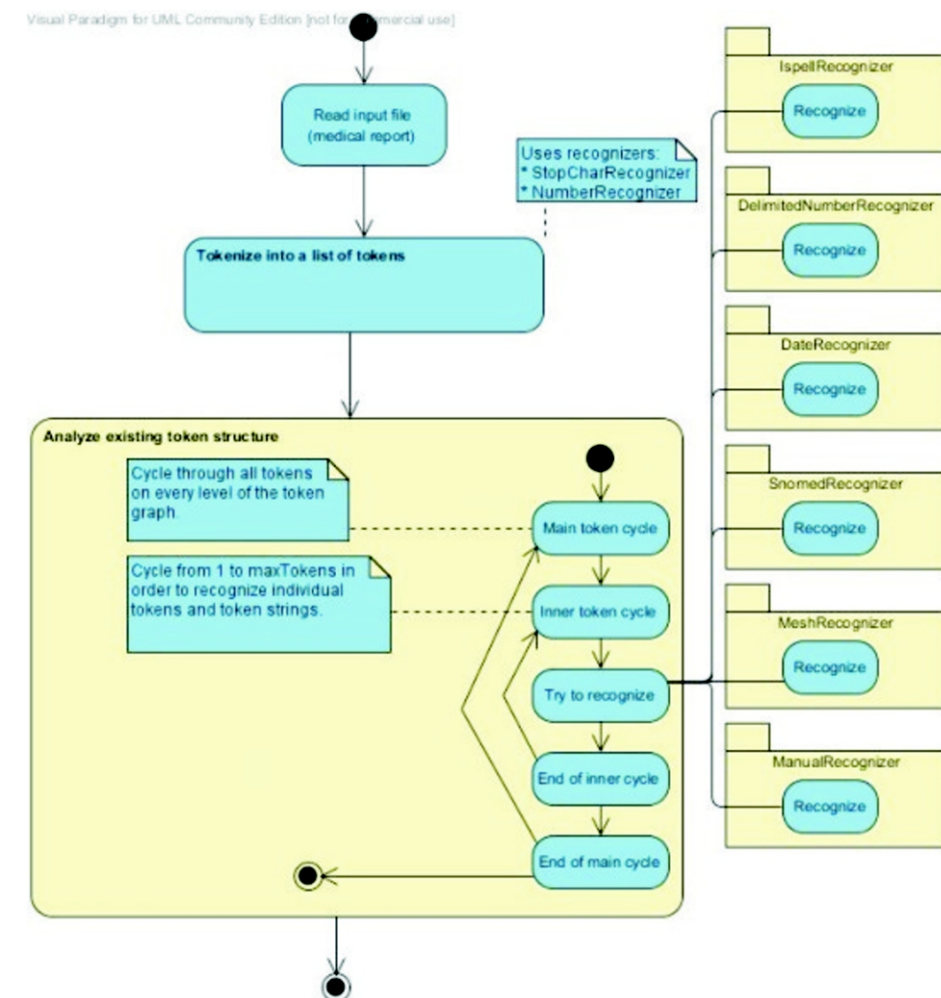
4. Identifikace číselných hodnot

Postup rozpoznávání popisuje Obrázek 1. V první části zpracování se zaměřujeme na úpravu textu, ve druhé části na využití konečných automatů pro rozpoznání definovaných struktur.

Úprava textu spočívá především v tokenizaci. Tokenizací vznikne seznam tokenů, ty jsou pro potřeby zpracujícího programu zapsány v obecných komponentách, které využíváme ve druhé části zpracování.

Například text „Krevní tlak:145/90“ přejde v následující seznam tokenů (kontejnerů), viz Tabulka 1.

Jednotlivé konečné automaty jsou implementovány jako třídy se společným rozhraním (interface); pracovně je označujeme „rozpознаваče“ (recognizers). Tyto třídy operují nad jedním nebo více kontejnerů. Kontejnery obsahují údaje o rozpoznaném podkladu. V této fázi používáme tyto typy kontejnerů:



Obr. 1. Schematický activity diagram běhu programu.

Date

Kontejner obsahuje rozpoznané datum. Datum rozpoznává třída *ManualRecognizer*. Podmínkou pro rozpoznání datumu jsou údaje o dni, měsíci a roce, oddělené tečkou a existence takového datumu (např. 32.12.2010 není rozpoznáno jako datum). Datum je rozpoznáváno z kontejneru typu *DelimitedNumber*.

DelimitedNumber

Obsahuje rozpoznaná čísla oddělená stejným druhem kontejneru *SpecialChar*. Tímto způsobem jsou k rozpoznání připravena data, rodná čísla či údaje o krevním tlaku.

Dictionary

Obsahuje rozpoznaná česká slova. Obsahuje klasifikaci slovního druhu. Tento

kontejner vytváří *IsSpellRecognizer*, založený na upraveném českém slovníku Ispell. Úprava českého slovníku Ispell souvisí s klasifikací slovního druhu.

MedicalTerm

Obsahuje rozpoznané položky klasifikačních systémů SNOMED CT a MeSH. Tento kontejner vytvářejí rozpoznávače *SnomedRecognizer* a *MeshRecognizer*. Tyto rozpoznávače jako podklad používají kontejnery základního typu (nerozpoznaná slova) a kontejnery typu *Dictionary*.

Tímto způsobem se velmi dobře daří identifikovat číslo pojištěnce a krevní tlak. Číslo pojištěnce v České republice obsahuje datum narození, informaci o pohlaví a rodná čísla přidělená v posledních přibližně 40 letech jsou tvořena tak, že musejí splňovat určitou ověřitelnou vlastnost. V případě zpráv, které máme k dispozici, se rodné číslo podařilo rozpoznat v přibližně 85 % případů.

Tab. 1. Seznam tokenů k textu „Krevní tlak: 145/90“.

Container „Krevní“	SpecialChar „ “ (1x)	Container „tlak“	SpecialChar „:“ (1x)	Number „145“	SpecialChar „/“ (1x)	Number „90“
-----------------------	-------------------------	---------------------	-------------------------	-----------------	-------------------------	----------------

Případy, kde nebylo rodné číslo rozpoznáno zahrnují především tyto situace:

- zpráva neobsahovala číslo pojištěnce (to je v rozporu s vyhláškou o zdravotnické dokumentaci) nebo
- zpráva obsahovala jen informaci o datu narození, v některých případech byl uveden jen rok narození nebo
- rodné číslo bylo zapsáno v nesprávném formátu (obvykle vynechaná nula ve druhé trojmístné části rodného čísla).

Struktura odpovídající tlaku byla správně rozpoznána v 96 % zpráv. Ve zbývajících případech nebyla informace o krevním tlaku identifikována, neboť krevní tlak nebyl zapsán ve formátu STK/DTK, ale zvlášť hodnota systolického tlaku a zvlášť hodnota diastolického tlaku. Pro rozpoznávání struktury krevního tlaku jsme využili běžného oddělovače složek této hodnoty (lomítko "/"), maximální přípustné uvažovatelné hodnoty systolického krevního tlaku a pravidla, že systolický krevní tlak je vždy vyšší než diastolický krevní tlak. Ve výjimečných případech byla jako krevní tlak označena hodnota, která krevní tlak nepředstavovala. Konkrétně šlo o popis parametrů EKG.

6. Identifikace termínů v lékařských zprávách

Pro identifikaci termínů jsme použili SNOMED CT ve verzi z roku 2008 [9], Mezinárodní klasifikaci nemocí verze 10 v české verzi distribuovanou ÚZIS [10], číselníky NČLP [11] a českou verzi MeSH [12].

Nejúspěšnější identifikace údajů z klasifikačních jsme podle očekávání dosáhli při použití české verze MeSH. MeSH totiž na rozdíl od české verze MKN10 téměř neobsahuje zkratky. Souhrnné údaje o zpracovaných souborech jsou uvedeny v Tabulce 2. Slova jsou v tomto případě označené tokeny zjištěné tokenizací. Mezi slovy tak jsou i samostatná čísla (např. údaj o krevním tlak "145/90" obsahuje dva tokeny považované za slova). Prvních 20 nejčastěji rozpoznaných českých deskriptorů klasifikace MeSH je uvedeno v Tabulce 3.

Termíny jsme se snažili identifikovat pomocí zjišťování shody částí textů (jednoho až pěti po sobě jdoucích slov) s hodnotami z číselníků.

Vyhledávací algoritmus vždy nejprve vyhledával termín odpovídající zjištěné kombinaci a v případě, že takový termín nenalezl, pokračoval vyhledáváním termínu, který hledaným textem pouze začínal. V případě, že hledaný text obsahoval tečku, algoritmus tečku nahradil za možnost, že slovo pokračuje libovolnými znaky. Přesto byla jak v případě SNOMED CT (anglická verze), tak v případě MKN10 úspěšnost identifikace velmi nízká.

V číselníku SNOMED CT se dařilo identifikovat termíny, které mají stejný název v angličtině i v češtině, obvykle jde o latinské termíny jako např. "diabetes mellitus" (SNOMED CT kód 73211009). Podobně se pomocí SNOMED CT dařilo identifikovat fyzikální jednotky (např. „MmHg“, SNOMED CT kód 259018001) a některé názvy léků z částí zpráv, které obsahovaly informaci o předepsané terapii. Neúspěch při identifikaci ostatních termínů jsme očekávali, neboť jejich český název neodpovídá anglickému.

Tab. 2. Souhrnné údaje o zpracovaných zprávách.

Počet zpráv	107
Počet slov (tokenů)	66.376
Počet rozpoznaných termínů (MeSH CZ)	1.126
Průměrný počet rozpoznaných termínů ve zprávě	10,52
Počet rozpoznaných termínů / počet slov (tokenů)	1,7 %

Tab. 3. Deset nejčastěji rozpoznaných českých deskriptorů.

Kód	Český deskriptor	Počet
H01.671.691	tlak	63
A07.541	srdce	62
C14.907.489	hypertenze	39
A01.047	břicho	39
C14.280.067	arytmie	37
A01.456	hlava	34
C23.550.288	nemoc	33
A01.598	krk	30
D04.808.247.222.284	cholesterol	29
E01.370.370.380.650	puls	28

Velkým zklamáním byla česká verze číselníku MKN10. ÚZIS (Ústav pro zdravotnické informace a statistiku ČR), která publikuje kompletní překlad Mezinárodní klasifikace nemocí, dostupný číselník však používá obrovské množství zkratk a zkrácených slov. Výše zmíněný diabetes mellitus uvádí kromě názvu "diabetes mellitus" také pod názvem "DM". Velké množství položek obsahuje více zkrácených slov, například položka T449 zní "j. a neurč.léč.působ.prim.na autonom .nerv.soustavu". Podobně u klasifikace stavů vzniklých z dopravních nehod jsou zkrácená slova velmi častá, např. pro kód V0602 zní "Chod.x.j.nemot.v.; neprovoz.n.; výdělečná čin." či pro kód V0920 zní "Chod.x zraněný při prov.neh.j.a neurč.mot.voz.". Uvedené dva příklady ukazují, že jedenkrát je slovo "vozidlo" zkráceno na "voz.", jindy jen na "v.", podobně "nehoda" někdy na "neh.", jindy na "n.". V případě, kdy zdravotní záznamy samotné obsahují značné množství zkrácených slov takový stav činí z české verze číselníku Mezinárodní klasifikace nemocí pro účely extrakce informací téměř nepoužitelný produkt.

6. Závěr

V současné době není kromě MeSH [12] k dispozici číselník s českými názvy diagnóz, který by bylo možné využít pro jejich identifikaci v lékařských zprávách.

Číselníky NČLP bohužel neobsahují seznam diagnóz a tak je nebylo možné použít pro jejich detekci. Jako užitečný se zdá být číselník jednotek NČLP, neboť jej lze využít pro vyhledání a klasifikaci jednotek. Číselníky NČLP bohužel nejsou označené registrovanými OID a nejsou ani indexovány UMLS [13].

Pokud by existovala česká verze SNO-MED CT, bylo by snáze možné rozpoznávat jednotlivé termíny. Možným kandidátem na slabší, nicméně využitelný nástroj, je klasifikace MeSH. Českou verzi klasifikačního systému MeSH vydává Česká lékařská knihovna a zpřístupňuje ji pomocí webového rozhraní a na základě individuální smlouvy o použití.

Výhodou MeSH je skutečnost, že je indexována v databázi UMLS a tak jsou jeho koncepty mapovatelné na klasifikační systémy použitelné ve strukturách standardů rodiny HL7 a EHRcom.

Tento článek popisuje současný stav naší práce. Ve snaze o extrakci informací z lékařských zpráv pomocí nestatistických metod budeme nadále pokračovat.

Poděkování

Práce byla částečně podpořena projektem 1M06014 MŠMT ČR a projektem SVV-2010-265513.

Literatura

- [1] Žďárek R.: Vedení zdravotnické dokumentace a její náležitosti. Zdravotnické noviny 30.6.2009
- [2] Hammond K., Helbig S., Benson C., Brathwaite-Sketoe B.: Are Electronic Medical Records Trustworthy? Observations on Copying, Pasting and Duplication. AMIA Annual Symposium Proceedings 2003; 269273.
- [3] Přechová P., Zvárová J.: The Role of International Nomenclatures and Standards in Travel Shared Health Care. Mihalas, G.; Saka, O.; Mazzoleni, T.; Blobel, B.; Pharow, P.; Gülkesen, K. eds. Travel Health Informatics and Telehealth. European Notes in Medical Informatics 5, 2009, 162-164
- [4] Safran Ch., Bloomrosen M., Hammond W., Labkoff S., Markel-Fox S., Tang P., Detmer D.: Toward a National framework for the Secondary Use of Health Data: An American

Medical Informatics Association White Paper. Journal Americal Medical Informatics Association 14 (1):2007, 1-9

- [5] Semecký J., Zvárová J. (školicitelka): Multi-mediální elektronický záznam o nemocném v kardiologii, Univerzita Karlova v Praze, Matematicko-fyzikální fakulta, Praha, 2001
- [6] Smatana P., Paralič J. (školicitel): Spracovanie lekárskych správ pre účely analýzy a dolovania v textoch, Technická univerzita v Košiciach, Košice, 2005
- [7] HaCohen-Kerner Y., Kass A., Peretz A.: Baseline methods for automatic disambiguation of abbreviations in Jewish law documents. Lecture Notes in Computer Science, Advances in Natural Language Processing Baseline methods for automatic disambiguation of abbreviations in Jewish law documents Vol 3230 Springer 2004, 58-69
- [8] Yu H., Hripcsak G., Friedman C.: Mapping Abbreviations to Full Forms in Biomedical Articles. , Journal of the American Medical Informatics Association 9 no.3, 2002, 262272
- [9] Systematized Nomenclature in Medicine – Clinical Terms, International Health Terminology Standards Development Organisation
- [10] Mezinárodní klasifikace nemocí, Ústav zdravotnických informací a statistiky / World Health Organization
- [11] Národní číselník laboratorních položek, Ministerstvo zdravotnictví ČR, <http://ciselniky.dasta.mzcr.cz>
- [12] Medical Subject Headings, Národní lékařská knihovna / United States National Library of Medicine
- [13] Unified Medical Language System, United States National Library of Medicine

Kontakt

Ing. Karel Zvára

Centrum biomedicínské informatiky
Ústav informatiky AV ČR, v.v.i.
Pod Vodárenskou věží 2
182 07 Praha 8
e-mail: karel@zvara.cz

Stav databáze Katalogu klinických doporučených postupů publikovaných prostřednictvím internetu v českém jazyce

M. Zvolský¹

¹Centrum biomedicínské informatiky, Ústav informatiky AV ČR v. v. i.

Školitel: Doc. Ing. Arnošt Veselý, Csc.

Souhrn

Od devadesátých let minulého století narůstá publikační aktivita odborných autorit při tvorbě dokumentů klinických doporučených postupů (KDP). Během první dekády 21. století v rámci trendu popularity sdílení informací v elektronické podobě se jedním z hlavních publikačních médií těchto dokumentů stává internet. Podle předchozího průzkumu webových sídel českých lékařských odborných společností a zkušeností z dosavadního provozu Katalogu klinických doporučených postupů publikuje v současné době na internetu přibližně 107 českých odborných společností až 731 KDP. Dokumenty klinických doporučených postupů jsou používány lékaři a ostatními klinickými pracovníky v klinické praxi, správními autoritami při posuzování a vyhodnocování konkrétních klinických postupů, tvůrci zdravotních politik, lékaři při vzdělávání, pacienti a v neposlední řadě jsou informace v těchto dokumentech obsažené klíčové pro tvorbu klinických informačních systémů, systémů pro podporu rozhodování a dalších aplikací medicínské informatiky.

Databáze Katalogu klinických doporučených postupů postupně od roku 2007 mapuje výskyt dokumentů KDP v rámci volně přístupných webových sídel českých odborných společností, případně odborných periodik těmito společnostmi garantovaných. V rámci této databáze jsou shromažďovány bibliografické informace, informace o aktuálnosti a kvalitě dokumentů a o existenci případných webových nástrojů a aplikací vytvořených na základě těchto KDP.

Databáze obsahuje 562 záznamů o dokumentech KDP publikovaných českými odbornými společnostmi. 212 záznamů je přístupných prostřednictvím webového rozhraní Katalogu klinických doporučených postupů.

Klíčová slova: internet, World Wide Web, databáze, klinický doporučený postup, klinická praxe, medicína založená na důkazech, formalizace, GLIF (Guideline Interchange Format), lékař, systémy pro podporu rozhodování

1. Úvod

1.1 Definice klinických doporučených postupů

Klinické doporučené postupy (KDP) jsou důležitou součástí rozhodovacího procesu na všech úrovních poskytování zdravotní péče [1]. Uznávanou definicí těchto dokumentů je, že se jedná o „systematicky vyvíjená sdělení, která mohou být využita k dosažení přiměřených specifických rozhodnutí, služeb a výsledků poskytované zdravotní péče“ [2].

Přes existenci velkého množství dokumentů v česky psaných odborných publikacích, které tuto definici v hrubých rysech splňují, bývá pro ně užito různých termínů, které se ne vždy plně překrývají, například: lékařské doporučení, doporučený postup, guideline, metodický pokyn, protokol, standard a další. Česká kardiologická společnost na svých webových stránkách používá „guideline“ i „doporučený postup“, Česká lékařská společnost Jana Evangelisty Purkyně pak na stejné úrovni používá pojmy „léčebný standard“ a „doporučený postup“. V poslední době však i v české literatuře dochází k jasnému vymezení používaných pojmů [1], [3], [4], [5].

1.2 Publikační aktivita odborných lékařských společností

Dokumenty definující zdravotnickou péči v konkrétních klinických souvislostech vznikají na různých úrovních hierarchie poskytování zdravotnické péče od publikací Světové zdravotnické organizace přes mezinárodní odborné společnosti,

Mezinárodní či národní normativy až po pouze lokálně platné směrnice jednotlivých poskytovatelů zdravotní péče.

V dalším textu budeme uvažovat dokumenty s platností na území České republiky, ať již vznikly zadáním a garancí odborné autority (nejčastěji lékařské odborné společnosti) nebo převzetím, či překladem dokumentu s mezinárodní či původně zahraniční platností garantovaným taktéž českou odbornou autoritou.

V českém právním řádu nejsou v současné době standardy zdravotní péče stanoveny obecně závaznými právními předpisy [6]. Přiměřenost zdravotní péče je hodnocena (například v případě pochybení příslušným kontrolním orgánem) jako „lege artis“ nebo „non lege artis“ nejen podle vztahu k existujícím klinickým doporučeným postupům zahrnujícím konkrétní klinický případ, ale také podle individuálního stavu pacienta a volba poskytované péče je vždy na rozhodnutí a odpovědnosti ošetřujícího lékaře.

V České republice jsou uznávanými odbornými autoritami pro publikování klinických doporučených postupů především odborné lékařské společnosti sdružené do České lékařské společnosti Jana Evangelisty Purkyně. K 1. 7. 2010 bylo těchto odborných společností 107 [5]. Systematickému vývoji dokumentů normativního charakteru určujících poskytování zdravotnické péče v daných klinických podmínkách se věnuje jen část z nich (viz Tabulka 1.) [7]. Dalšími potenciálními publikačními autoritami na tomto poli jsou Česká lékařská komora, Česká stomatologická komora, Ministerstvo zdravotnictví ČR, případně Státní zdravotní ústav či Hlavní hygienik ČR.

Tab. 1. Počty dokumentů KDP publikovaných elektronicky prostřednictvím webových sídel českých lékařských odborných společností nebo jejich oficiálních periodik, rok 2009 (uvedeny pouze společnosti s osmi a více dokumenty).

Název odborné společnosti	Počet dokumentů
Společnost všeobecného lékařství	47
Česká kardiologická společnost	24
Česká dermatovenerologická společnost	36
Česká gastroenterologická společnost	14
Česká neurologická společnost	18
Česká pneumologická a fteologická společnost	22
Česká angiologická společnost	8
Česká diabetologická společnost	13
Česká hepatologická společnost	18
Česká onkologická společnost	79
Česká revmatologická společnost	9
Česká společnost klinické biochemie	34
Česká společnost pro aterosklerózu	10
Společnost infekčního lékařství	19
Společnost urgentní medicíny a medicíny katastrof	17
Společnost českých patologů	9
Společnost pro transfuzní lékařství	8
Radiologická společnost	13
Česká hematologická společnost	14
Česká společnost anesteziologie, resuscitace a intenzivní medicíny	13
Česká lékařská společnost Jana Evangelisty Purkyně	305

Dokumenty KDP jsou publikovány a šířeny několika způsoby. V tištěné verzi KDP vycházejí jako články v odborných periodikách lékařských odborných společností (např. Cor et Vasa, Česká kardiologická společnost), jednak vznikají samostatné brožury (např. edice Doporučený diagnostický a léčebný postup pro všeobecné praktické lékaře, Centrum doporučených postupů pro praktické lékaře, Společnost všeobecného lékařství).

V elektronické verzi vycházejí KDP buď jako články v elektronických verzích odborných časopisů lékařských společností (např. Vnitřní lékařství, Česká internistická společnost) [8] nebo častěji jako dokumenty umístěné na webových

prezentacích jednotlivých lékařských společností (např. Česká dermatovenerologická společnost) [9]. Ve všech případech se jedná o elektronickou verzi textových lékařských doporučení, nejčastěji ve formátech Portable Document Format (PDF), Hypertext Markup Language (HTML) nebo Microsoft Office Document (DOC) [7].

Protože odborné společnosti vyvíjejí KDP samostatně, nepadá shoda o podobě těchto dokumentů a jednotlivé KDP se od sebe liší nejen po formální stránce, ale také co do množství obsažených informací. Výjimkou jsou dokumenty Společnosti všeobecného lékařství, která pro tvorbu těchto dokumentů ustanovila Centrum doporučených postupů, vyvinula

základní metodiku a dokumenty vydává v ucelené edici brožur dostupných i v elektronické podobě.

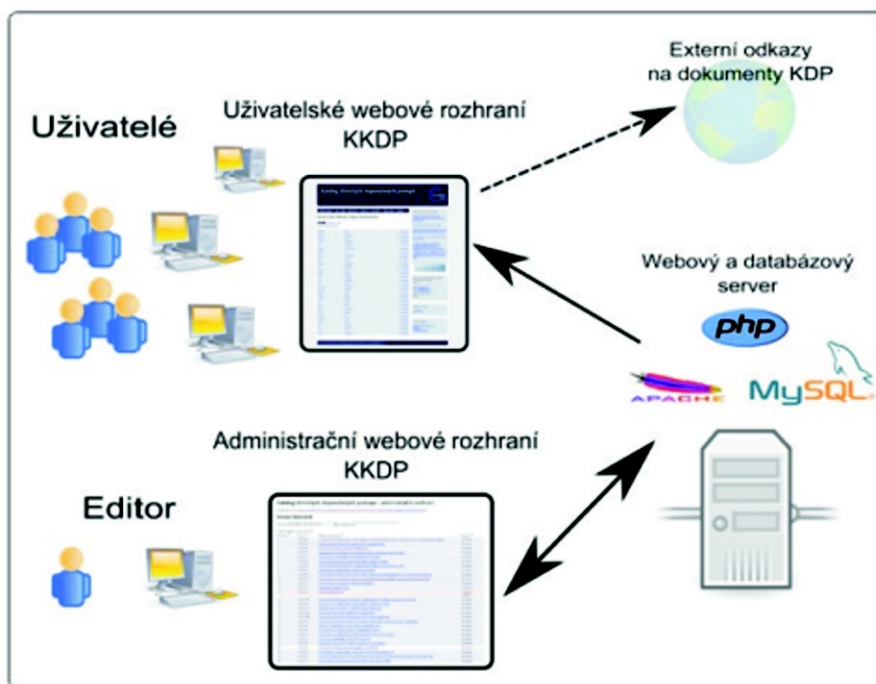
V souvislosti se snahou o zavedení standardů léčebné péče jako striktně stanovené součásti zdravotnického systému v České republice vzniká také Národní sada standardů zdravotních služeb vyvíjená Národním referenčním centrem. Pro vývoj těchto standardů byla vypracována poměrně podrobná metodika. Jednotlivé standardy však prozatím nejsou veřejně dostupné [9].

1.3 Katalog klinických doporučených postupů – historie a cíle projektu

Velké množství dokumentů KDP a jejich rozdílná forma, množství publikujících subjektů a také velké množství webových sídel, kde jsou tyto dokumenty v elektronické formě umístěny, způsobuje problémy při vyhledávání konkrétního dokumentu pro konkrétní klinickou situaci.

Při použití fulltextových internetových vyhledávacích služeb nebo obecných katalogů se přidává také problém kvality vyhledaných dokumentů. Texty zmiňující klinické situace a jejich řešení mohou publikovat zdravotnická zařízení, zpravodajské portály věnující se problematice zdraví, pacienti, zájmová sdružení, farmaceutické firmy a další subjekty. Takové dokumenty jsou určeny různým cílovým skupinám, mohou se v odborné kvalitě velmi lišit od skutečně platných KDP nebo jim dokonce po obsahové stránce odporovat.

V zahraničí existují specializované portály věnující se problematice KDP. Jejich součástí jsou i katalogy elektronicky publikovaných KDP. Nejznámějšími katalogy KDP jsou National Guidelines Clearinghouse v USA [10] a National Library of Guidelines ve Velké Británii [11], obojí zaměřené na anglicky psané dokumenty. Také v dalších státech vznikají národně orientované katalogy (např. <http://www.leitlinien.de/>, <http://www.nzgg.org.nz/>). Národní organizace do mezinárodní sítě s katalogem dokumentů spojuje Guidelines International Network (G-I-N). [12]



Obr. 1. Katalog klinických doporučených postupů.

Přes členství českého Národního referenčního centra v G-I-N nejsou doposud v žádném zahraničním či mezinárodním katalogu informace o českých dokumentech KDP. Žádný ze zahraničních katalogů také nedisponuje webovým rozhraním v českém jazyce, které by umožnilo snadnější přístup českých uživatelů bez znalosti anglického jazyka.

Od roku 2007 vyvíjíme na našem pracovišti Katalog klinických doporučených postupů (KKDP), dříve také pod názvem Lékařské doporučené postupy v ČR. [13] Podnětem pro jeho vznik byla absence přehledného seznamu dokumentů KDP v českém jazyce (resp. platných na území České republiky) publikovaných elektronicky prostřednictvím internetu. Inspirací byla existence zahraničních webových katalogů stejného určení. Vzniklo webové rozhraní programované v jazyce PHP, které umožňuje procházení záznamů uložených v databázi MySQL a jejich řazení podle různých kritérií. Toto veřejně přístupné uživatelské webové rozhraní je umístěno na adrese <http://neo.euromise.cz/kkdp>.

Druhé webové rozhraní (administrační) přístupné po zadání přihlašovacích údajů a určené pro správce a editory systému umožňuje editaci a přidávání jednotlivých

záznamů komunikaci s autory dokumentů KDP i mezi editory, kontrolu platnosti

externích odkazů a další funkce. Schematická struktura projektu KKDP je znázorněna na Obrázku 1.

Databáze nashromážděných údajů o jednotlivých dokumentech KDP má ovšem širší uplatnění. Vzhledem ke shromažďování informací o autorech a odborných autoritách publikujících KDP, může sloužit pro adresář těchto osob a institucí. Do budoucna se počítá také s návazností jednotlivých dokumentů KDP na odkazy na relevantní publikace odkazy například do databáze Medline. Databáze je připravená na navázání informací o formálních modelech formalizovaných verzích KDP, modely však v současné době nejsou pro české KDP dostupné. Díky informacím o formálních modelech a provázání na lasifikační systémy a tezaury je možným využitím zdroj informací pro systémy na podporu rozhodování [14]. Budoucí rozšíření databáze se předpokládá s rozvojem sociálních a komunitních funkcí portálu KKDP.

Tab. 2. Struktura databázové tabulky _documents.

Název sloupce	Datový typ	Popis
category_id	int(11)	kategorie dokumentu KDP
document_abstract	text	abstrakt dokumentu KDP
document_actualization	date	datum poslední aktualizace dokumentu KDP
document_authornote	text	poznámka k autorům dokumentu KDP
document_contact	text	kontaktní informace na autory dokumentu KDP
document_creation	date	datum vytvoření/schválení dokumentu KDP
document_id	int(11)	pořadové číslo dokumentu KDP (webové rozhraní)
document_iid	varchar(20)	kódové označení dokumentu KDP
document_jinda	datetime	datum a čas vložení informací o dokumentu KDP (administrační rozhraní)
document_keywords	text	klíčová slova dokumentu KDP
document_lcda	datetime	datum a čas poslední změny informací o dokumentu KDP (administrační rozhraní)
document_lcui	smallint(6)	identifikace uživatele, který naposledy změnil záznam (administrační rozhraní)
document_note	text	textová poznámka k dokumentu KDP
document_printed	text	úplný název tištěné verze dokumentu KDP
document_state	int(11)	stav záznamu z hlediska ověření informací (administrační rozhraní) – určuje, zda je záznam plně ověřen editorem a má se zobrazit v uživatelském rozhraní
document_target_population	text	cílová populace pro dokument KDP jako celek
document_termination	date	datum ukončení platnosti dokumentu KDP
document_title	text	název dokumentu KDP
document_usid	smallint(6)	identifikace uživatele, který vložil záznam (administrační rozhraní)
document_views	int(11)	počet shlédnutí detailu o dokumentu KDP (webové rozhraní)
evidence_id	int(11)	nejnižší použitá úroveň důkazů podle Oxford Centre for Evidence-Based Medicine
document_evidence	text	definice a popis úrovně důkazů použitých v dokumentu KDP
recency_id	smallint(5)	typ aktuálnosti dokumentu KDP – určuje, zda se jedná o návrh dokumentu, platný dokument nebo již obsolentní dokument

2. Metodika

Při tvorbě databáze pro KKDP jsme vycházeli jednak ze zásad Conference on Guideline Standardization (COGS) [15], jednak ze vzoru katalogu National Guideline Clearinghouse [10] s velmi vyčerpávajícími informacemi o každém dokumentu KDP v tzv. „Guideline Summary“. Tyto vzory byly ovšem konfrontovány s reálnou situací, kdy samotné dokumenty KDP publikované českými autoritami nejen nesplňují požadavky COGS, často u nich nejsou uvedeny ani základní identifikační údaje jako autor nebo doba platnosti dokumentu. Zároveň nebylo možné využít komunikaci s autory dokumentů či s jinými odborníky a informace doplnit. Ačkoliv se v projektu KKDP počítá s intenzivní komunikací a spoluprací odborníků a autorů KDP ke spoluvytváření katalogu a doplňování chybějících údajů, nebyly a nejsou dosud k dispozici prostředky ani zdaleka pro české měřítko srovnatelné s možnostmi např. Ame-rické Agency for Health-care Research and Quality, která provozuje National Guideline Clearinghouse. V první fázi bylo naší snahou zaznamenat informace v dokumentech KDP již obsažené, případně snadno dohledatelné (například související onemocnění v číselníku Mezinárodní klasifikace nemocí 10. revize). Proto byl kladen důraz na strukturované identifikační údaje o tvůrcích, časovém a odborném vymezení KDP. Další informace je možné doplnit například do univerzálního textového sloupce `document_note` v Tabulce `_documents`.

Současná verze 2.0 Katalogu klinických doporučených postupů vzniká od začátku roku 2010. Pro uložení záznamů je použita databáze MySQL ve verzi 5.0.41. Databáze obsahuje celkem 49 tabulek. 18 tabulek obsahuje záznamy o dokumentech, autoritách, autorech, kódy použitých slovníků a klasifikací a pomocné číselníky, 13 tabulek obsahuje data potřebná pro funkci webových rozhraní a 18 tabulek je vazebních.

K dokumentu KDP mohou být v KKDP přiřazeny některé kódy mezinárodních klasifikačních a nomenklaturních systémů. Jsou to:

Tab. 3. Struktura databázové tabulky `_linked_files`.

Název sloupce	Datový typ	Popis
<code>file_id</code>	<code>int(11)</code>	pořadové číslo záznamu odkazu
<code>file_name</code>	<code>text</code>	název odkazu
<code>file_link</code>	<code>text</code>	znění hypertextového odkazu ve formě URL
<code>filetype_id</code>	<code>smallint(6)</code>	typ souboru (reference to the external code list)
<code>file_note</code>	<code>text</code>	poznámka k odkazu
<code>file_size</code>	<code>int(11)</code>	velikost souboru v kilobytech
<code>file_inda</code>	<code>datetime</code>	čas vložení záznamu o odkazu (administrační rozhraní)
<code>file_usid</code>	<code>int(11)</code>	identifikace uživatele, který vložil záznam (administrační rozhraní)
<code>file_lcda</code>	<code>datetime</code>	čas poslední změny záznamu o odkazu (administrační rozhraní)
<code>file_lcui</code>	<code>int(11)</code>	identifikace uživatele, který poslední změnil záznam (administrační rozhraní)
<code>file_status</code>	<code>smallint(6)</code>	typ aktuálnosti dokumentu KDP – určuje, zda se jedná o návrh dokumentu, platný dokument nebo již obsolentní dokument
<code>file_show</code>	<code>boolean</code>	určuje, zda bude odkaz zobrazen v uživatelském rozhraní
<code>file_actualization</code>	<code>date</code>	datum poslední aktualizace odkazovaného souboru
<code>broken_link</code>	<code>boolean</code>	označení nefunkčního linku automatickou kontrolou

1. *Medical Subject Headings (MeSH)* – komplexní tezaurus pro biologické vědy
Každému dokumentu KDP lze přiřadit jeden nebo více kódů určujících pozici deskriptoru (klíčového slova) v hierarchickém stromu. Jeden deskriptor se může v hierarchickém stromě MeSH vyskytovat opakovaně na různých místech, v tom případě je vhodné použít kód, který nejlépe vystihuje způsob použití deskriptoru v dokumentu KDP.

2. *Mezinárodní klasifikace nemocí (MKN), 10. revize*

Každému dokumentu KDP lze přiřadit jeden nebo více troj- až čtyřpísmenných kódů této klasifikace.

3. *Diagnosis-related groups (DRG) systém klasifikace klinických případů*
Každému dokumentu KDP lze přiřadit jeden nebo více pětimístných číselných kódů této klasifikace.

4. *Snomed CT (Systematized Nomenclature of Medicine – Clinical Terms) – systém zdravotnické terminologie*
Každému dokumentu KDP lze přiřadit jeden nebo více konceptů tohoto systému. Snomed CT prozatím nemá český překlad a žádný dokument KDP proto vymezení konceptů Snomed CT neobsahuje. Vhodné koncepty lze dohledat po přeložení klíčových termínů do angličtiny nebo nama-

pováním výše zmíněných nomenklatur či klasifikací na Snomed CT.

Kódy klasifikačních systémů jsou ke každému dokumentu KDP přiřazeny editorem v administračním rozhraní KKDP. Každému dokumentu je možné přiřadit libovolný počet souvisejících kódů. Buď jsou obsaženy ve výčtu v samostatné části dokumentu nebo je nutné je aktivně vyhledat v plném textu dokumentu. Ve většině případů však samotný text neobsahuje přímo kódy klasifikačních systémů. V tom případě je nutné vyhledat v textu klíčové termíny a k nim pomocí externích nástrojů vyhledat patřičné kódy.

Katalog klinických doporučených postupů používá i další číselníky jako:

- Seznam autorit (především KDP publikujících odborných společností) je průběžně doplňován při zadávání nových dokumentů KDP.
- Seznam autorů KDP – je průběžně doplňován při zadávání nových dokumentů KDP.
- Seznam lékařských a zdravotnických specializací podle IPVZ.
- Číselníky odkazující na geografickou příslušnost nebo jazykovou verzi dokumentu.

Obr. 2. Snímek obrazovky generovaný webovým prohlížečem z adresy http://neo.euromise.cz/kkdp/?akce=mkn_strom_ukazuje_prochazeni_hierarchickou_strukturou_klasifikace_mkn_s_odkazy_na_definice_mkn_kodu_na_webovych_strankach_ustavu_zdravotnickych_informaci_a_statistiky_cr.

- Dokument je volně přístupný, zdarma, bez registrace, není chráněn heslem ani jiným způsobem ochrany.

Vyhledávání dokumentů probíhá dvojím způsobem. Jednak jsou od roku 2007 procházena webová sídla odborných lékařských společností (podle seznamu umístěného na stránkách České lékařské společnosti Jana Evangelisty Purkyně) a ručně jsou zařazovány do katalogu dokumenty, pokud splňují výše uvedená kritéria. Druhým možným způsobem je zařazování na základě tipů uživatelů katalogu, kteří mohou ze stránek KKDP zaslat elektronický formulář s vytipovaným dokumentem, který doposud nebyl zařazen.

Po vytvoření nového záznamu KDP je k dokumentu vyhledáno maximální možné množství informací, které jsou prostřednictvím administračního rozhraní editorem/administrátorem zadány.

Hlavní tabulka `_documents` obsahuje 23 sloupců definujících dokument KDP (viz Tabulka 2). Další číselníky jsou připojeny prostřednictvím vazebních tabulek.

Stejným způsobem je připojena i tabulka externích odkazů `_linked_files`, která ke každému dokumentu definuje jeden či více odkazů, tedy umístění dokumentu nebo jeho částí v prostředí internetu. Tato tabulka obsahuje dalších 14 sloupců blíže definujících daný odkaz (viz Tabulka 3).

Pro zařazení dokumentu do KKDP jsou určena následující kritéria:

- Dokument je dostupný v elektronické formě prostřednictvím internetu.
- Dokument je garantován odbornou autoritou, nejčastěji odbornou lékařskou společností (viz Úvod), je přístupný z webových stránek této autority nebo jejího oficiálního periodika a je umístěn na stejném webovém sídle.
- Dokument je definován ve svém názvu nebo souvisejícím popisu na webových stránkách odborné autority jako guideline, doporučení, doporučený postup, standard, metodický pokyn nebo protokol.

Obr. 3. Snímek obrazovky generovaný webovým prohlížečem z adresy <http://neo.euromise.cz/kkdp/?akce=detail&dokument=46> - Detail informací o dokumentu klinického doporučeného postupu s odkazy na plné změny on-line.

Do budoucna se počítá s komunikací a účinnou spoluprací přímo s autory KDP, kteří nashromážděné informace potvrdí nebo doplní. U každého záznamu je možné určit fázi zpracování, která vymezuje, zda jsou informace o dokumentu vhodné k zobrazení v uživatelském rozhraní katalogu z hlediska kompletnosti a aktuálnosti.

Pro udržování aktuálnosti katalogu byl do administračního rozhraní zařazen nástroj na automatické procházení a vyhodnocení existence nebo neexistence dokumentu na základě kontroly HTTP hlavičky s adresou daného odkazu. Pokud hlavička HTTP vrací chybový kód, je tento odkaz označen jako chybný a zobrazen editorovi/administrátorovi, který ověří, zda se jedná o krátkodobý výpadek nebo se dokument na daném umístění již nevyskytuje, pokusí se ho dohledat, informaci o dokumentu v katalogu upravit nebo dokument z katalogu vyřadit.

Uživatelské rozhraní KKDP poskytuje seznam záznamů o dokumentech KDP řazený podle názvu či data poslední změny. Dalšími možnostmi jsou řazení podle autorů (abecedně) nebo podle autorit (odborných společností). Pro vyhledání vhodného dokumentu podle tematické příslušnosti lze využít hierarchický strom MeSH nebo hierarchický strom MKN (viz Obr. 2). Ze seznamu dokumentů podle libovolného řazení nebo vyhledávání lze vybrat a zobrazit detailní výpis informací o dokumentu (viz Obr. 3), ve kterém jsou vypsány všechny dostupné informace o daném dokumentu včetně hypertextových odkazů na on-line přístupné plné znění textu KDP.

3. Výsledky

K 1. září 2010 obsahovala databáze Katalogu klinických doporučených postupů 562 záznamů o dokumentech KDP. 12 těchto záznamů bylo označeno jako odstraněný/neexistující. 212 záznamů bylo označeno jako ověřený a bylo zobrazováno prostřednictvím uživatelského webového rozhraní KKDP.

Databáze obsahovala 640 záznamů o odkazovaných souborech. 17 z nich bylo označeno jako chybný odkaz.

Databáze ke stejnému dni obsahovala 491 záznamů o autorech dokumentů a 91 záznamů o odborných autoritách.

V uživatelském rozhraní se zobrazují pouze dokumenty ve stavu ověřený a autoři a odborné autority zúčastněné alespoň na jednom dokumentu ve stavu ověřený. Proto jsou v KKDP veřejně přístupné pouze informace o 212 dokumentech, 430 autorech a 46 autoritách.

4. Diskuse

Databáze v současné podobě mapuje většinu současné publikační aktivity v oblasti KDP dostupných prostřednictvím Internetu. Celkový počet těchto dokumentů je minimálně 731 odkazovaných souborů [7], v databázi je aktuálně zaznamenáno 640, resp. 623 z nich.

V porovnání s některými zahraničními katalogy klinických doporučených postupů jako Guidelines International Network, National Guideline Clearinghouse nebo National Library of Guidelines [12], [10], [11] obsahuje databáze našeho katalogu menší množství záznamů. Je to způsobeno především zaměřením našeho katalogu na česky psané dokumenty. Ve srovnání s počtem anglicky publikovaných dokumentů česká odborná literatura zaostává. Ne všechny záznamy ve zmíněných zahraničních katalozích obsahují také odkaz na on-line volně přístupný plný text doporučeného postupu. Srovnání se zahraničními katalogy s národním zaměřením nebo omezené na jednu organizaci například New Zealand Guidelines Group [16] nebo Scottish Intercollegiate Guidelines Network [17] vyznívá v počtu

Katalogizovaných dokumentů příznivěji (viz Tabulka 4). Na rozdíl od zahraničních webových katalogů má náš katalog větší možnosti řazení dokumentů podle různých parametrů.

Mnohé z českých dokumentů KDP jsou natolik stručné nebo neobsahují dostatek identifikačních údajů, že je nutné informace vhodné pro zařazení do KKDP pracně dohledávat a doplňovat. Na toto úsilí chybějí prostředky, proto v současnosti katalog zobrazuje z 550 aktuálních záznamů pouze 212, u nichž jsou informace vloženy dostatečně reprezentativní.

Dokumenty KDP většinou neobsahují kódy terminologických slovníků nebo klasifikací, obsahují často textové informace o nosologických jednotkách, patogenech nebo např. anatomických strukturách, které lze ve slovnících dohledat. Toto dohledávání je méně přesné, než kdyby kódy byly určeny přímo v textu nebo doplněny autorem. K nepřesnostem může docházet především ve chvíli, kdy se popisovaný termín přesně neshoduje s textem, je vymezen širší nebo úžeji, nebo se nachází v hierarchickém stromě nomenklaturního systému na více místech. Databáze KKDP nabízí možnost zadání tří v České republice běžně používaných klasifikací a slovníků (MeSH, MKN, DRG) a slovníku SNOMED CT, který prozatím nemá český překlad, ale je nejkomplexnější. Zadávání všech čtyř slovníků/klasifikací naráží na problémy nedostatku informací v původním textu a také na to, že termíny všech čtyř si někdy plně neodpovídají.

Tab. 4. Srovnání některých webových katalogů dokumentů klinických doporučených postupů.

Jméno katalogu	Počet dokumentů	Řazení podle
Katalog klinických doporučených postupů	623 (212 ověřených)	autor, organizace, MeSH, MKN
Guidelines International Network	7100	MeSH
National Guideline Clearinghouse	2549	MeSH, organizace
National Library of Guidelines	3220	specializace
New Zealand Guidelines Group	74	léčebná kategorie, organizace
Scottish Intercollegiate Guidelines Network	120	téma

Při správném pokrytí a vyplnění těchto informací u každého dokumentu KDP (nejlépe ve spolupráci s autorem dokumentu, který by zařazení provedl nejpřesněji) by se zvyšovaly možnosti napojení KKDP na zdravotnické informační systémy (MKN, DRG, v budoucnu SNOMED CT) i bibliografické systémy (často používaný MeSH).

Databáze by měla do budoucna obsahovat také informace o formalizovaných modelech dokumentů KDP a to ve formě odkazů na popis těchto modelů nebo odkazů na webové aplikace, která tyto modely využívají nebo zobrazují. V současné době databáze tyto informace neobsahuje především z toho důvodu, že formální modely KDP neexistují nebo nejsou prostřednictvím internetu veřejně přístupné. V budoucnosti by však KKDP mohl sloužit i jako on-line repozitář informací o těchto modelech.

Databáze KKDP je uzpůsobena pro uchovávání informací o dokumentech KDP v českém jazyce. Pro české prostředí ovšem mohou být platné dokumenty s geograficky širší působností, například dokumenty vydané Světovou zdravotnickou organizací nebo mezinárodními odbornými společnostmi. Tyto dokumenty mohou existovat také v cizojazyčných verzích. Proto v budoucím vývoji bude databáze KKDP uzpůsobena pro podporu více jazyků.

Další plánovanou úpravou databáze je vytvoření možnosti přidávat k dokumentům KDP informaci o relevantní literatuře s provázáním například do bibliografické databáze Medline. Databáze KKDP bude i v jiných ohledech dále upravována podle potřeb projektu.

Cílem projektu je nadále udržovat informace o dokumentech KDP se snahou o úplné pokrytí publikačních aktivit odborných autorit. Proto budou mimo jiné dále

upravovány nástroje pro zautomatizování procesu kontroly platnosti odkazů a nově také pro automatické procházení webových sídel odborných autorit a tipování nových dokumentů k zařazení. Cílem je také využití komunikace s autory dokumentů a spolupráce s nimi, případně vytvoření odborné komunity kolem projektu KKDP.

5. Závěry

Za účelem shromažďování informací o dokumentech klinických doporučených postupů s platností pro Českou republiku je od roku 2007 vyvíjen Katalog klinických doporučených postupů. Součástí tohoto projektu je i databáze záznamů o dokumentech KDP ve formátu MySQL, která může mít do budoucna další využití v systémech pro podporu rozhodování. K 1. září 2010 obsahovala databáze 550 aktivních záznamů o dokumentech KDP na celkem 623 platných umístěních. Databáze bude dále rozšiřována a přizpůsobována publikačním aktivitám odborných autorit.

Poděkování

KKDP je vyvíjen s podporou projektu 1M06014 MŠMT ČR.

Tento výstup vznikl v rámci projektu SVV-2010- 265 513.

Literatura

- [1] Ličeník R.: Klinické doporučené postupy. Výukový portál LF UP Olomouc [online] 20.11.2009, poslední aktualizace 20.11.2009 [cit. 2010-09-24], dostupný z: <http://mefanet.upol.cz/clanky.php?aid=38>
- [2] Field M.J., Lohr K.N.: Guidelines for Clinical Practice: From Development to Use, NATIONAL ACADEMY PRESS, 1992
- [3] Suchý M.: Příručka pro autory a oponenty vývoje klinických doporučených postupů, Národní referenční centrum [online] 1. 9. 2010, dostupný z: http://www.sopr.cz/standards/dokumenty/dokument_open.php?doc=3
- [4] Česká kardiologická společnost, <http://www.kardio.cz>

- [5] Česká lékařská společnost Jana Evangelisty Purkyně, <http://www.cls.cz>
- [6] Těšínová J.: Právní aspekty poskytování zdravotní péče v ČR, Medical Tribune 18/2010, dostupný z: <http://www.tribune.cz/clanek/18816-pravni-aspekty-poskytovani-zdravotni-pece-v-cr>
- [7] Zvolský M.: Základní parametry dokumentů doporučených postupů českých lékařských společností publikovaných prostřednictvím Internetu, Doktorandské dny'09, MATFYZPRESS; 2009, dostupný z: <http://www.cs.cas.cz/hakl/doktorandskyden/files/2009/sbornik-dd-2009.pdf>
- [8] Vnitřní lékařství, <http://www.vnitrnilekarstvi.cz>
- [9] Národní referenční centrum, <http://www.nrc.cz>
- [10] National Guideline Clearinghouse, <http://www.guidelines.gov>
- [11] National Library of Guidelines, <http://www.library.nhs.uk/guidelinesfinder>
- [12] Guidelines International Network, <http://www.g-i-n.net>
- [13] Zvolský M.: Katalog lékařských doporučených postupů v ČR, Doktorandské dny'08, MATFYZPRESS; 2008, dostupný z: <http://www.cs.cas.cz/hakl/doktorandskyden/files/2008/sbornik-dd-2008.pdf>
- [14] Buchtela D., Peleska J., Zvolský M., Zvarova J.: Medical knowledge representation system, Stud Health Technol Inform. 2008;136:377-82.
- [15] Shiffman R.N., Shekelle P., Overhage J.M., Slutsky J., Grimshaw J., Deshpande A.M.: Standardized reporting of clinical practice guidelines: a proposal from the Conference on Guideline Standardization. Ann Intern Med. 2003 Sep 16;139(6):493-8
- [16] New Zealand Guidelines Group, <http://www.nzgg.org.nz>
- [17] Scottish Intercollegiate Guidelines Network, <http://www.sign.ac.uk>

Kontakt

MUDr. Miroslav Zvolský

Centrum biomedicínské informatiky
Ústav informatiky AV ČR v. v. i.
Pod Vodárenskou věží 2
182 07 Praha 8
e-mail: zvolsky@euromise.cz