Medical Informatics Europe '99
P. Kokol et al. (Eds.)
IOS Press, 1999

Automated computer-assisted evaluation of diagnosis-and-procedure-reports in Austrian hospitals

Hagenbichler*, E.; Klingler** , D., Neuner** , L., Pfeiffer, K.-P.***
** Jörgerstr. 3/35, A-1170 Wien.; ** AKH Linz, A-4020 Linz;*
**** Inst. f. Biostatistik, Univ.-Klinik Innsbruck, A-6020 Innsbruck*

Abstract. Austria's 320 hospitals with 73.000 beds have to transmit a minimum basic data set for each hospital stay with diagnosis and procedures as well as scores for use in intensive care units (last only in performance-oriented financed hospitals). Based on these data, 160 hospitals or 53.000 beds are financed through performance oriented flat rates per case (LDFs - Leistungsorientierte Diagnosen-Fallpauschalen). Those LDFs are in some principles similar to DRG´s (diagnosis related groups), but have been developed by the Federal Ministry of Labor, Health and Social Affairs (BMAGS). 1.94 million data sets of hospital stays, which were reported for financing in the year 1997 had to be analyzed for their correctness and plausibility. For this aim the "Scoringprogramm" of the BMAGS is in use. We describe in detail how many diagnoses and procedures are automatically checked with diagnosis, sex, age, length of stay etc. 1.6% warnings resulted from this checking, and no errors have been reported. The checking routines have to be further developed. At present data seem to be used only with caution for epidemiology or for analyzing medical quality. For financing purposes it might be reliable, especially if data quality commissions are enlarged and if evaluation of data with medical records is improved.

1. Background

Austrian hospitals are obliged to transmit an anonymous Minimum Basic Data Set (MBDS) for each hospital stay to the provincial funds, which is then forwarded to the Federal Ministry of Labor, Health and Social Affairs (BMAGS) [1]. It is a must that one main diagnosis and up to 9 additional diagnoses are being documented among other information like age, sex, ward, admission and discharge dates and administrative data. Since 1999 there is the possibility to gather the diagnosis voluntarily on the basis of ICD-10 (International Statistical Classification of Diseases and related health problems, 10^{th} Revision, WHO – World Health Organization) [2]. This diagnosis will then be automatically encoded into the (mandatory) Austrian version of ICD-9-code numbers for the reason of hospital financing [3]. For this purpose an unequivocal translation table has been provided by the BMAGS [4].

Furthermore hospitals are obliged to document up to 9 medical procedures including their quantities. The documentation of medical procedures is based on an Austrian-specific catalog of medical procedures. In 1999 this catalog comes with 925 medical procedures and with 8 scores for therapeutic-intervention and severity-of-illness, whereas the 8 scores are

used for the documentation of the performance in intensive care units (ICUs) only in performance-oriented financed hospitals.

Since 1998, these 8 scores have to be documented since 1998 using a different data set [5,6]. At admission one of the following scores has to be documented – depending on the patient's age: either SAPS II (Simplified Acute Physiology Score) [6,7], TRISS (a combination of Trauma Score and Injury Severity Score) [6,8], CRIB (Clinical Risk Index for Babies) [6,9], SNAP (Score for Neonatal Acute Physiology) [6,10] or PRISMIII (Pediatric Risk of Mortality Score) [6,11]. For daily documentation the following are in use: TISS-28 (Therapeutic Intervention Scoring System) [6,12], NTISS (Neonatal TISS) [6,13] or PedTISS (pediatric TISS) [6,14].

The correct use and hints for the software implementation and for plausibility checks of these classifications, catalogs and scores are described in a handbook and in additional publications of the BMAGS [3,6,15,16,17,18,19]. In the year 1997 there were earlier versions of the Austrian version of the ICD-9 and of the catalog of procedures in use [20,21].

Based upon this documentation and through performance-oriented flat rates per case (LDFs - Leistungsorientierte Diagnosen-Fallpauschalen) - that in some principles are similar to diagnosis related groups (DRG) - (but which we have especially developed in Austria), the scores (which represent the costs) have been calculated for the hospitals all over Austria in the same way since 1997. Days on ICUs are paid in addition. The paid cost for one day on a specific ICU is mainly determined through the average of TISS-points per day of the year before.

Some private hospitals and trauma surgery hospitals are exceptions of this financing system. For instance, in 1997 about 20.000 beds (roughly 160 smaller hospitals) were excepted and 53.000 beds (about 160 hospitals) were included in this financing system [22]. Before 1997 and in the private and trauma surgery hospitals nowadays a daily flat-rate was or respectively is paid.

1.94 million data sets of hospital stays (one for each stay), which were reported for financing in the year 1997 had to be put to the proof for their completeness and plausibility.

In the following we present the software tool for the year 1997 and for 1999, which we have developed for the routine check and the plausibility, and which is available free of charge on the internet in its actual version (http://www.bmags.gv.at, Scoringprogramm Version 3.1) [23]. We will also describe how these goals have been reached and which further checking routines will have to be developed.

As the new Austrian performance-oriented hospital financing system (LKF – Leistungsorientierte Krankenanstaltenfinanzierung) [24] is based upon these data, the correctness of these data is not only of big importance from an epidemiological point of view, but also from an economic one!

2. Results

We investigated the checking routines for the plausibility as defined by the BMAGS, which are mandatory in use and which are represented as software tools, namely "Scoringprogramm Version1.2" for the year 1997 [25] and "Version 3.1" for the year 1999 [23]; "Intensivscoringprogramm Version 2.1" [26] is used for the collection of data in ICUs for the year 1999.

The scoring program serves for calculating the points for financing the hospital stay of the patient, for checking the plausibility of data, and, if necessary, also for encoding the diagnosis and procedures in the hospital.

The scoring program is implemented under Windows 3.1. A C-source-code is available. Free disk space less than 6 MB is needed for the program itself (3 MB of the 6 MB are needed for documentation files). With the program the BMAGS - as the overall responsible organization for data quality - checks the plausibility of the data, which takes approximately one working day when using a standard PC for checking ca. 2 million data sets of one year [22].

The program of 1999 for the data of ICUs (under Windows 3.1; 4 MB disk space needed) checks completeness, the lab results in relation to medical minimum/maximum human limits and also for the combination of various parameters of the scores, e.g. "treatment of metabolic alkalosis or acidosis" with the respective "pH-value".

On the one hand the scoring program of 1997 evaluates formal criteria of completeness. On the other hand there are different other criteria. First diagnoses are checked in relation to age. 516 diagnoses out of 6.151 diagnoses are checked with age restrictions, e.g. "ICD-9 - 647.8 – toxoplasmosis during pregnancy" with age limits from 5 to 60 years, or "ICD-9 - 769 – IRDS (infants respiratory distress syndrome)" is checked with age limits from 0 to 4, or "ICD-9 - 797 – senility" with limits from 65 to 150. In 1999 there are 520 out of 6774 diagnoses with age restrictions.

For 1997 there are 628 diagnoses related to sex. Examples are "ICD-9 - 174 - malignant neoplasm of the mamma" with sex female, or "ICD-9 – 600 - hyperplasia of the prostate" with sex male. In 1999 there are 661 sex restrictions.

Rare and / or notifiable forms of disease are also checked in 71 cases in 1997 (in 1999: 64 cases), e.g. "ICD-9 - 050 – smallpox", or "ICD-9 – 001 - cholera".

A large part of all medical procedures (as of 1997: out of a catalog with 812 medical procedures [20], in 1999 with a total of 925 medical procedures [19]) is evaluated in connection to diagnosis (1997: 14.216 relations, 1999: 18.568 relations). The diagnosis is cross-checked only with the first three digits of the ICD-9-code. Thus in 1997, 767 diagnoses out of 937 three digits codes were checked as being plausible with 598 procedures out of the above mentioned 812. The median of one diagnosis is checking with 11 procedures (minimum 1, maximum 298), whereas the median checking of one procedure is with 13 diagnoses (minimum 1, maximum 146). In 1999 there are 858 diagnoses out of 937 three digits codes checked as plausible with 697 procedures out of the above mentioned 925. The median of one diagnosis is checking with 12 procedures (minimum 1, maximum 351), whereas the median checking of one procedure is with 16 diagnoses (minimum 1, maximum 232).

Furthermore checking is being done in respect to the number of performed medical procedures (all procedures with an upper high and very high limit). For example a coding of a "laparascopic appendectomy" or a "cholecystectomy" more often than one time during one hospital stay will cause a warning (upper high limit), but if it is coded more than four times it is doubtlessly an error (very high limit). In these cases, one can discuss if it makes sense to increase the very high limit to more than 4, and indeed a reduction of this limit has been done to 3 and probably will be done to 1 in the future.

In 1997 the checking of procedures in connection with male/female sex of patients had 69 relations, in 1999 78 relations. Examples are "hysterectomy" with female or "operation of a phimosis" with male patients.

In 1997 the checking of procedures with age occurred 17 times, in 1999 20 times, e.g. "Caesarian operation" only for patients aging between 5 and 60; or some age-pre-defined operations like "operation of the diaphragm in children under the age of 1 month".

Since 1999, 126 diagnoses are checked also if they are the only plausible cause for hospitalization. Maybe these patients should be rather treated as out-patients in the clinic

than as in-patients in the ward, and the same may be true with superficial wounds, and some other specific symptoms.

823 diagnoses are checked if they are approved of as a cause for a "0-day-hospital-stay" (i.e. admission and normal discharge within the same day). Here the diagnosis seems to be too serious for being treated as 0-day-hospital-stay, e.g. "coma" or "sepsis".

Since 1999 medical procedures are also checked for their plausibility if performed within a "0-day-hospital-stay" and normal discharge (538 procedures, e.g. "heart transplantation" and other major operations, "therapy within a stroke unit", etc.) or if they could be rather handled as out-patients, such as with "transesophageal echocardiography" or maybe also with "chronic hemodialysis" (19 relations).

"Post-mortem examination" is cross-checked if it fits to the type of discharge (dead).

The generated warning and error hints usually necessitate to check the data with the medical records.

If the check for age, sex and the higher limit of the number of medical procedures produces an "error" result, this hospital stay of the respective patient will not be financed until the error is corrected. The other checking routines produce a "warning" result, so that a control with the medical record is necessary.

Very commonly occurring warnings concerned the combination of diagnosis with the procedure, e.g. "removal of osteosynthesis material" with the previous diagnosis of the acute bone fracture instead of the correct diagnosis, the delayed consequence of this fracture. Other warnings resulted from a misunderstanding of the definition of a procedure as for example in "continuous measurement of the intracranial pressure", where many physicians counted this "per day" instead of "per the whole time while continuously measuring" [22].

After an introduction phase we realized, however, that in some cases, the restrictions were too rigorous, so the checking routines needed adaptation. We also realized that there were many procedures which were not clearly defined, so we had to define a new and more precise description and counting of procedures.

Indeed in some cases also the choice of the correct diagnosis as main diagnosis was changed, there had been some misunderstandings with the coding hospital physician as well as with the administrative hospital employees. Also combinations of medical procedures have been coded, which were an implicit part of the main procedure, e.g. a therapeutic arthroscopy where also a diagnostic arthroscopy during the same procedure has been coded (because one has to look into the patient). Another example is the coding of two tonsillectomies during one operation because of the argument that both tonsils had been removed.

In 1997 there were approximately 1.6 % automatically produced warnings (ca. 32.000 of approx. 1,940.000 transmitted data sets) [22]. No errors have been reported [22].

The results of the evaluation are reported to the provincial funds (so-called "Landesfonds"), and they usually give feedback to the hospitals, although the hospitals are obliged to make this evaluation by themselves.

The provincial funds have established groups or commissions for checking the data plausibility ("data quality groups/commissions"). At the current state, these groups are working more or less intensively. Before the introduction of the LKF-system the various health insurance organizations took a bigger influence in the evaluation of the medical treatment, especially concerning the length of stay and allowance for treatment as in-patient.

3. Further developments

Future evaluation criteria should relate the diagnosis with the type of medical department, and therefore evaluate the correct choice of the main diagnosis. Further criteria could be a linking of all hospital stays of a patient in the same hospital to check the type of admission.

Further checking routines could be extended to the principle of severity-of-illness of diagnoses (including their combination). Also the severity-of-illness of one or more medical procedures and their relation with length of stay, type of department, type of admission and discharge (e.g. death) could be further developed. The analysis of combination of procedures should be of great importance, and therefore it would be necessary to know the date of the procedure (which we do not know at present).

Last but not least, an analysis of statistical relations in collectives of patients can be made regarding to type of hospital (e.g. in relation to the points that represent costs). A tool for producing statistically valid samples could be useful. A relation to other official reports should be made (e.g. carcinoma register or mortality statistics), allowing further evaluation of data.

In the future one should put more attention to the evaluation and interpretation of diagnosis-and-procedure-reports as tools for planning and information provision in the public health sector. Therefore methods which are now in use in the field of artificial intelligence (e.g. knowledge-based expert systems, data-mining and neural networks) should be further developed and implemented for this purpose.

4. Conclusion

The software, currently available free of charge, covers a lot of questions for automatic evaluation of data and their plausibility to choose suspicious data sets. After checking the medical records, these data sets have to be corrected if necessary.

However, the current quality and quantity of data is not sufficient to give detailed information on questions of medical quality and of epidemiology. But the data can be quite helpful for financing purposes and if one accepts a relatively broad range of acuity for spreading the financial means.

But even for this purpose the evaluation of data has to be increased through the work of the data quality groups or commissions.

The random sampling of patients data sets and checking them with the medical records will always be necessary.

References

[1] http://www.ris.bka.gv.at Dokumentation im Gesundheitswesen. BGBl Nr. 745/1996 and 783/1996.
[2] http://www.bmags.gv.at ICD-10 BMAGS 1999
[3] http://www.bmags.gv.at ICD-9 BMAGS 1998
[4] http://www.bmags.gv.at Überleitungstabelle ICD-10 nach ICD-9
[5] http://www.ris.bka.gv.at Dokumentation im Gesundheitswesen. BGBl II Nr. 63/1998 and 473/1998
[6] http://www.bmags.gv.at Medizinisches Dokumentationshandbuch (inkl. 2 Anhänge)
[7] SAPS-II: Le Gall JR, Lemeshow St, Saulnier F. A new Simplified Acute Physiology Score (SAPS II) based on a European/North American Multicenter Study. JAMA 1993; 270(24): 2957-2963.
[8] TRISS: Boyd CR, Tolson MA, Copes WS. Evaluating Trauma Care: The TRISS Method. The Journal of Trauma. April 1987.
[9] CRIB: The International Neonatal Network. The CRIB (clinical risk index for babies) score: a tool for assessing initial neonatal risk and comparing performance of neonatal intensive care units. Lancet 1993; 342: 193-198.
[10] SNAP: Richardson, Douglas K., et al. Score for Neonatal Acute Physiology: A Physiologic Severity Index for Neonatal Intensive Care. Pediatrics. 1993;91:617-623
[11] PRISMIII: Pollack, Murray M, et al. PRISM III: An updated Pediatric Risk of Mortality score. Crit Care Med. 1996;24: 743-752.
[12] TISS-28: Miranda DR, De Rijk A, Schaufeli W. Simplified Therapeutic Interventions Scoring System: The TISS-28 items - Results from a multicenter study. Crit-Care-Med 1996; 24(1): 64-73.
[13] NTISS: Gray, James E., et al. Neonatal Therapeutic Intervention Scoring System: A Therapy-Based Severity-of-Illness Index. Pediatrics. 1992;90: 561-567.
[14] PedTISS: Pollack, Murray M, Capron, Carolanne. The Pediatric Risk of Mortality and the Therapeutic Intervention Scoring System. In: Essentials of Pediatric Intensive Care. Levin, Daniel L., Quality Medical Pub. 1990; 31-37
[15] http://www.bmags.gv.at Satzartenbeschreibung + Corrigenda
[16] http://www.bmags.gv.at 1. LKF-Intensiv-Rundschreiben
[17] http://www.bmags.gv.at 2. LKF-Intensiv-Rundschreiben
[18] http://www.bmags.gv.at 7. LKF-Rundschreiben
[19] http://www.bmags.gv.at MEL-Katalog BMAGS 1999.
[20] ICD-9-BMGK 1997. BMAGS Gruppe VII/A, Stubenring 1, A-1010 Vienna.
[21] MEL-Katalog BMGK1997. Katalog medizinischer Einzelleistungen.
[22] Information by BMAGS, Gruppe VII/A, Stubenring 1, A-1010 Vienna.
[23] http://www.bmags.gv.at Scoringprogramm V.3.1
[24] http://www.ris.bka.gv.at Vereinbarung gemäß Art. 15a B-VG über die Reform im Gesundheitswesen und der Krankenanstalten. BGBl I Nr. 111/1997
[25] Scoringprogramm V.1.2. für 1997. BMAGS, Gruppe VII/A, Stubenring 1, A-1010 Vienna.
[26] Intensiv-Scoringprogramm V.2.1. für 1999. BMAGS, Gruppe VII/A, Stubenring 1, A-1010 Vienna.