Exploratory Analysis of Medical Coding Practices: the Relevance of Reported Data Quality in Obstetrics-Gynaecology

Diana LUNGEANU\textsuperscript{a1}, Daniela ZAHARIE\textsuperscript{b}, Stefan HOLBAN\textsuperscript{c}, Elena BERNAD\textsuperscript{d}, Maria BARI\textsuperscript{e}, Rodica NOAGHIU\textsuperscript{a}

\textsuperscript{a}Department of Medical Informatics, University of Medicine and Pharmacy of Timisoara, Romania
\textsuperscript{b}Faculty of Mathematics and Computer Science, West University of Timisoara, Romania
\textsuperscript{c}Faculty of Computer Science, Polytechnic University of Timisoara, Romania
\textsuperscript{d}Department of Obstetrics-Gynaecology, University of Medicine and Pharmacy of Timisoara, Romania
\textsuperscript{e}Department of Obstetrics-Gynaecology, University of Medicine and Pharmacy of Bucharest, Romania

Abstract. We aimed at identifying a suitable data analysis approach to investigate potential patterns in the current medical coding in obstetrics and perinatal care. We processed the data reported for 2006 in DRG files from three Romanian university clinics of obstetrics-gynaecology and found substantial differences in the coding practices. Based on the evidence we found with a poor usage of the coding instruments, we concluded that using objective methods and quantifiable measures in analyzing the medical coding could help putting things into the right perspective and bring support for the need for formal education of medical record administrators and coders where such programmes do not exist, e.g. in Romania.

Keywords. Data analysis tools, coding, diagnosis related, outcomes research and measurement

Introduction

The Diagnosis Related Groups (DRG) system was initially developed as an information management tool to monitor quality and use of services but is now widely used as a prospective payment system in most European countries \cite{1, 2}. Although comprising data collected mainly at the discharge moment and typically used for measuring the severity of illness, so resulting in the case-mix index for each specific hospital as a basis for financing and/or reimbursement, the reported DRG files also contain valuable information for quality management within the hospital sector \cite{3, 4}:

\footnotesize\textsuperscript{1} Corresponding Author: Diana Lungeanu, Department of Medical Informatics, University of Medicine and Pharmacy, P-\textsuperscript{ta} E. Murgu 2, 300041 Timisoara, Romania; E-mail: dlungeanu@umft.ro
main and secondary diagnoses (the latter important as co-morbidities) based on the International Classification of Diseases v10 or ICD-10 [5, 6], procedures’ codes, length of stay, etc.

As in 2005 the DRG system became compulsory for the hospital care in Romania, we tried to use this reported minimum data set to perform a patient profiling in obstetrical and perinatal care. However, in processing the data we encountered quality and reliability problems (data initially collected from a single hospital), so we decided to further investigate the medical coding patterns, as data validity and reliability are important criteria in choosing appropriate sources and measurements for assessing healthcare services [2]. The DRG system is at the beginning in Romania and the code assignment affects the case-mix and the financing in this prospective payment system. It should thus not be surprising that the hospitals try to “accommodate” the code usage to their perceived needs for financing, trying to balance this trade-off against the correct coding from a medical point of view. This has been a sensitive problem wherever the system was introduced [1, 7].

Taking this into account, we decided to perform an analysis of the coding practices in obstetrics and evaluate the use of codes, aiming at discovering the factors which influence the current practices.

### 1. Methods Used in Investigating Medical Coding Practices

Our goal was to investigate the coding results in hospitals of obstetrics-gynaecology in order to determine whether there were any coding patterns or “common practices” (i.e. good vs. bad habits). Empirical data comprising the ICD-10 codes used for the principal and secondary diagnoses (i.e. the main medical problem and the comorbidities or complications) were imported from the DRG application compulsorily used in Romanian hospitals for reimbursement purposes.

We combined basic statistical analysis for the available ICD-10 codes’ usage with data mining techniques. We used SPSS v.15 and Weka 3-4 to process the data.

#### 1.1. Statistics Used to Characterize the Medical Coding Practice

The ICD-10 utilization statistics consisted of: (i) number of secondary codes (reflecting co-morbidities) recorded in the DRG files; (ii) percentage of codes used, of those specific for each specialty (obstetrics-gynaecology and perinatal care, respectively); (iii) number of codes covering up to 50% of cases for the principal and secondary diagnoses; (iv) a concentration index borrowed from industrial economics and introduced to healthcare by Spangler et al [4].

The degree of concentration across the coding categories was measured using the Hirschman-Herfindahl index (HHI), defined as it follows in Eq. (1).

\[
HHI = \sum_{i=1}^{n} p_i^2
\]

The value range for HHI is between 0 and 10000: the larger the number, the higher the degree of codes’ concentration (i.e. a small number of recorded/reported codes account...
for a disproportionately large number of all medical diagnoses or problems, co-
morbidities, etc. in that specialty and/or hospital). An HHI number below 1000 is
considered to indicate a fair degree of concentration for medical codes’ usage [4].

1.2. Clustering Methods Used to Explore the Medical Coding Practice

The aim of a data clustering task is to identify homogeneous groups of similar data
which satisfy at least the following conditions: (i) data in each group are sufficiently
similar; (ii) data in different groups are sufficiently dissimilar. Our investigation started
from a concern regarding the validity and reliability of the reported data set, so we tried
to see whether the observed cases were “parcelled” across the available ICD-10 codes,
i.e. we were looking for patterns of diagnostic codes’ utilization. We used the actual
codes (as nominal variables) and the length of stay (as a numerical variable, illustrating
the severity of the episode of care), expecting to find natural clusters comprising
similar clinical cases, on the condition that a reliable and consistent manner of coding
was used with respect to the World Health Organization’s recommendations [5, 6].
Moreover, we assumed that for equivalent teaching hospitals with similar pathologies
one can expect to obtain similar clustering results. We tried the K-means and
Expectation Maximization (EM) methods.

The K-means algorithm [8] is a partitioning approach based on the idea of
minimizing the average distance between the data in a cluster and the cluster’s center
(i.e. the prototype). It consists of an iterative process which starts from some randomly
selected centers, one for each cluster to be discovered. It can be easily adapted to
categorical data by replacing the average of data in a cluster with the element which
has a maximal similarity with all the other elements of the cluster (the so-called
medoid), so we considered it as appropriate for our data set with both nominal and
numerical variables.

In contrast with K-means where the clusters are disjoint, the EM algorithm
proposed by Dempster [9] relies on a statistical interpretation of the clustering
problems: data are considered as a mixture of probability distributions and the
technique seeks to identify this mixture’s parameters. Similarly to K-means, EM is also
based on an iterative process involving two main steps at each iteration: (i) expectation
step, which computes the probabilities of data with respect to the current values of the
mixture parameters; (ii) maximization step, which estimates the new parameters of the
probability distributions corresponding to clusters, based on a maximum likelihood
approach. This algorithm allowed us a statistical approach in determining the optimum
number of clusters for each data set: (a) we first applied the EM algorithm on each
dataset in order to estimate the appropriate numbers of clusters; (b) those values were
further used to compute the clusters’ prototypes by the K-means technique.

2. Results and Discussion

We analyzed data for 2006 in DRG files from three university hospitals, one in a
distant geographic location from the other two: HB (6862 cases), HC (5831 cases), HO
(10301 cases). In this paper, we focus on the exploratory methods for evaluating ICD-
10 codes’ usage. We also tried to get a perspective on the medical coding in Romanian
hospitals and to identify the potential patterns in the current medical coding practices.
Table 1 presents the statistical analysis. When calculating the codes’ usage statistics for new-born babies (NB), we considered as specific the 408 codes from classes P00-P96 (Certain conditions originating in the perinatal period) and Z30-Z39 (Persons encountering health services in circumstances related to reproduction). Similarly, when calculating the codes’ usage statistics for OG, we considered as specific the 715 codes from classes C51-C58 (Malignant neoplasms – Female genital organs), D06-D07 (In situ neoplasms – Carcinoma in situ), D25-D28 (Benign neoplasms), D39 (Neoplasms of uncertain or unknown behaviour), N70-N77 (Inflammatory diseases of female pelvic organs), N80-N98 (Noninflammatory disorders of female genital tract), O00-O99 (Pregnancy, childbirth and the puerperium), and Z30-Z39 (Persons encountering health services in circumstances related to reproduction).

At first, when comparing medical coding patterns, we might think it would be difficult to compare HHI for different specialties for they had completely different characteristics (e.g. intensive care compared to radiology or hematology). However, as the statistics prove in Table 1, even in the specific case of one single specialty (i.e. obstetrics-gynaecology or neonatal care) there is a high degree of heterogeneity across the three hospitals, as well as within the same hospital for the two different specialties.

Table 1. Statistics regarding the ICD-10 codes’ usage in the three hospitals (HB, HC, HO) for the wards of new-born babies (NB) and obstetrics-gynaecology (OG). HHI stands for Hirschman-Herfindahl index.

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>HB</th>
<th>HC</th>
<th>HO</th>
</tr>
</thead>
<tbody>
<tr>
<td>No of cases</td>
<td>2197</td>
<td>4665</td>
<td>2390</td>
</tr>
<tr>
<td>Length of stay (in days)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>min—max</td>
<td>0—31</td>
<td>0—34</td>
<td>0—369</td>
</tr>
<tr>
<td>mean</td>
<td>4.65</td>
<td>4.83</td>
<td>9.39</td>
</tr>
<tr>
<td>(std dev)</td>
<td>(2.20)</td>
<td>(3.26)</td>
<td>(32.21)</td>
</tr>
<tr>
<td>No of secondary codes/case</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>min—max</td>
<td>0—8</td>
<td>0—10</td>
<td>1—18</td>
</tr>
<tr>
<td>mean</td>
<td>0.69</td>
<td>1.66</td>
<td>3.69</td>
</tr>
<tr>
<td>(std dev)</td>
<td>(0.73)</td>
<td>(2.53)</td>
<td>(1.47)</td>
</tr>
<tr>
<td>* % of ICD-10 codes used for principal diagnosis</td>
<td>24%</td>
<td>24%</td>
<td>11%</td>
</tr>
<tr>
<td>* No of principal codes covering 50% of cases</td>
<td>8</td>
<td>3</td>
<td>7</td>
</tr>
<tr>
<td>* HHI for principal diagnosis codes</td>
<td>609</td>
<td>1293</td>
<td>1500</td>
</tr>
<tr>
<td>No of secondary codes covering 50% of cases</td>
<td>3</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>HHI for secondary codes</td>
<td>966</td>
<td>3113</td>
<td>1853</td>
</tr>
</tbody>
</table>

* Codes from the classes A00-B99 (Certain infectious and parasitic diseases), D50-D89 (Diseases of the blood and blood-forming organs and certain disorder involving the immune mechanism), Q00-Q99 (Congenital malformations, deformations and chromosomal abnormalities), and R00-R99 (Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified) were not included among those specific for female genital, maternity care or perinatal period, therefore not included in the denominator of any calculated ratios and percents. When encountered in the files, they had very low frequencies, so we considered them as outliers and completely excluded them from the codes’ usage statistics (i.e. included neither in the numerator nor the denominator).

** One hundred six cases had as principal diagnosis D18.0 (Haemangioma, any site from the class of Benign neoplasms) and were excluded from the statistics regarding the ICD-10 usage for new-born babies.
Although all three clinics are of similar size and of the same specialty, significant differences were obtained with respect to the number of clusters found by the EM method: 8 (HB), 16 (HC), and 6 (HO). Table 2 presents the prototypes we identified by the K-means method for one of the hospitals. The codes that appear in a cluster’s prototype (consisting of the mean values for the numerical attribute and the mode for the categorical ones) could be interpreted as best-describing the instances (i.e. cases) composing that group. Looking at the codes’ pattern for HB, clusters 2, 3, and 6 are identical, the only difference being in the length of stay. We chose to use length of stay as a clustering attribute, for it is one of the criteria influencing the DRG codes and the case-mix. Surprisingly the length of stay is not reflected in the coding pattern, though the longer the stay, the more complex should have been the case and that extra-burden of the disease should have been reflected in the secondary diagnoses. This flatness or lack of detail might be due to a rather poor coding, which is also suggested by the HHI for OG (Table 1 - high values for both the principal and the secondary codes). On the other hand, a good aspect is reflected by cluster 0, with D25.0 as the principal diagnosis code, correctly corresponding to the patients with the main diagnosis in the Neoplasms class (C00-D48). However, there still is a problem with this cluster, as the secondary codes are related to postpartum examination/care and to the babies’ care, so conflicting with the principal code. It also seems the code Z24.6 (specific to the new born babies) “engulfed” all the secondary codes for the OG clusters (for which the HHI is high as well).

Table 2. Clusters for HB identified by the K-means algorithm (8 clusters, 13 attributes, 6832 instances). For each cluster, the specialty, the number of instances and the percent are specified. Although all the available secondary diagnoses were considered in the clustering process, only the first four are presented in the table.

<table>
<thead>
<tr>
<th>Cluster centroids</th>
<th>Length of stay (mean±sd)</th>
<th>Princ diag</th>
<th>Sec diag1</th>
<th>Sec diag2</th>
<th>Sec diag3</th>
<th>Sec diag4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cluster 0 OG</td>
<td>446 (6%)</td>
<td>9.58±4.92</td>
<td>D25.0</td>
<td>Z39.2</td>
<td>Z24.6</td>
<td>Z24.6</td>
</tr>
<tr>
<td>Cluster 1 OG</td>
<td>1025 (15%)</td>
<td>6.82±2.69</td>
<td>O82.0</td>
<td>Z39.2</td>
<td>Z24.6</td>
<td>Z24.6</td>
</tr>
<tr>
<td>Cluster 2 OG</td>
<td>1136 (17%)</td>
<td>4.35±2.69</td>
<td>O80.0</td>
<td>Z39.2</td>
<td>Z24.6</td>
<td>Z24.6</td>
</tr>
<tr>
<td>Cluster 3 OG</td>
<td>1754 (26%)</td>
<td>2.28±0.78</td>
<td>O80.0</td>
<td>Z39.2</td>
<td>Z24.6</td>
<td>Z24.6</td>
</tr>
<tr>
<td>Cluster 4 NB</td>
<td>1037 (15%)</td>
<td>5.75±2.50</td>
<td>P59.9</td>
<td>Z39.2</td>
<td>Z24.6</td>
<td>Z24.6</td>
</tr>
<tr>
<td>Cluster 5 NB</td>
<td>207 (3%)</td>
<td>4.73±1.65</td>
<td>P21.1</td>
<td>Z24.6</td>
<td>Z23.2</td>
<td>P59.9</td>
</tr>
<tr>
<td>Cluster 6 OG</td>
<td>304 (4%)</td>
<td>7.65±2.39</td>
<td>O80.0</td>
<td>Z39.2</td>
<td>Z24.6</td>
<td>Z24.6</td>
</tr>
<tr>
<td>Cluster 7 NB</td>
<td>953 (14%)</td>
<td>3.44±0.99</td>
<td>Z38.0</td>
<td>Z39.2</td>
<td>Z24.6</td>
<td>Z24.6</td>
</tr>
</tbody>
</table>

The titles (i.e. medical meaning) for the ICD-10 codes appearing in Table 2 and/or in the text are: D25.0 Benign neoplasm – Submucous leiomyoma of uterus; O80.0 Spontaneous vertex delivery; O82.0 Single delivery by elective caesarian section; P21.1 Mild and moderate birth asphyxia; P59.9 Neonatal jaundice (physiological, intense, prolonged), unspecified; Z24.6 Need for immunization against viral hepatitis; Z38.0 Singleton, born in hospital; Z39.2 Postpartum care and examination - Routine postpartum follow-up

The concrete results for the other two hospitals were completely different from HB, i.e. different patterns corresponding to the clusters. On the other hand, they were similar in the sense of their intriguing clusters’ centroids, in a consistent manner with the HHI values (i.e. the higher HHI values, the meaningless the clusters).
3. Conclusions

Using exploratory analysis techniques, we investigated the medical coding in three university hospitals that offer top-quality medical care. We employed basic statistics regarding the specific codes’ usage, a concentration index borrowed from economics, and clustering techniques from data mining.

The conclusions are consistent with the initial perspective: there is a poor usage of the medical coding instruments. Based on these results and on the fact that two of the hospitals were located in the same geographic area (i.e. with similar patient profiles), we are confident the heterogeneities and difficulties are mainly due to the inexperience in applying the constraints of the DRG system. However, poor data quality leads to unreliable performance indicators and prevents appropriate and timely decisions, entailing subsequent costs. Therefore, we do think that using objective methods and quantifiable measures in analyzing the medical coding practices could help putting things into the right perspective and eventually obtaining good quality medical data.

Our study underpins the need for formal education programmes for medical record administrators (e.g. university programmes) to prepare professionals in charge with supervising medical coding activities in hospitals. In Romania we do not have such professionals and, moreover, most often the coders are the medical doctors themselves.

Acknowledgements

We appreciate the valuable contribution of Dr. Dan Navolan from the University Hospital of Obstetrics-Gynaecology “Dr. D. Popescu” from Timisoara. We also kindly thank Professors Ioana Moisil and Bernard Richards for their encouragement and useful comments on the early results of our work.

The study was supported by grant 99-II CEEX 03 - INFOSOC 4091/31.07.2006 from the Romanian Ministry of Education and Research.

References