Toward the Measure of Credibility of Hospital Administrative Datasets in the Context of DRG Classification

Diana Pimenta[1], Julio Souza[1,2](✉), Ismael Caballero[3], and Alberto Freitas[1,2]

[1] MEDCIDS – Department of Community Medicine, Information and Health Decision Sciences, Faculty of Medicine, University of Porto, Alameda Prof. Hernâni Monteiro, 4200-319 Porto, Portugal
juliobsouza@gmail.com
[2] CINTESIS – Center for Health Technology and Services Research, R. Dr. Plácido da Costa, 4200-450 Porto, Portugal
[3] University of Castilla-La Mancha, Ciudad Real, Spain

Abstract. Poor quality of coded clinical data in hospital administrative databases may negatively affect decision making, clinical and health care services research and billing. In this paper, we assessed the level of credibility of a nationwide Portuguese inpatient database concerning the codification of pneumonia, with a special emphasis on identifying suspicious cases of upcoding affecting proper APR-DRG (All-Patient Refined Diagnosis-Related Groups) classification and hospital funding. Using data on pneumonia-related hospitalizations from 2015, we compared six hospitals with similar complexity regarding the frequency of all pneumonia-related diagnosis codes in order to identify codes that were significantly overreported in a given facility relatively to its peers. To verify whether the discrepant codes could be related to upcoding, we built Support Vector Machine (SVM) models to simulate the APR-DRG system and assess its response to each discrepant code. Findings demonstrate that hospitals significantly differed in coding six pneumonia conditions, with five of them playing a major role in increasing APR-DRG complexity, being thus suspicious cases of upcoding. However, those comprised a minority of cases and the overall credibility concerning upcoding of pneumonia was above 99% for all evaluated hospitals. Our findings can not only be relevant for planning future audit processes by signalizing errors impacting APR-DRG classification, but also for discussing credibility of administrative data, keeping in mind their impact on hospital financing. Hence, the main contribution of this paper is a reproducible method that can be employed to monitor the credibility and to promote data quality management in administrative databases.

Keywords: Data quality · All Patient-Refined Diagnosis-Related Groups · Clinical coding · Hospital administration · Data quality management · Data governance

M. Piattini et al. (Eds.): QUATIC 2019, CCIS 1010, pp. 289–296, 2019.
https://doi.org/10.1007/978-3-030-29238-6_21

1 Introduction

The Diagnosis Related Groups (DRGs) classification system is currently employed in several countries worldwide and was developed to group hospital cases into specific clusters of patients (DRGs) with similar resource use and costs [1, 2]. The DRG system heavily relies on the quality of the data held in administrative databases, mainly regarding standard codes representing diagnoses and inpatient procedures [3]. Portugal's hospital financing system currently uses a more refined version of the DRG system, the All Patient-Refined Diagnosis-Related Groups (APR-DRG) version 31 [4]. From the point of view of data, the APR-DRG system can be understood as Master Data Repository, where the data model should be aligned to a standardized data model or vocabulary, and the collected values corresponding to codes of medical diagnoses and procedures should be aligned to a set of reference data. Codes representing the principal (hospitalization cause) or secondary (additional) diagnoses and inpatient procedures are captured by using the corresponding standard code from the ICD-9-CM (The International Classification of Diseases, Ninth Revision, Clinical Modification) classification system. Each APR-DRG represents the patient's reason for hospital admission, either in terms of disease or procedure performed [5].

In Portugal, data used for APR-DRG grouping purposes is acquired from hospital administrative databases, which were originally extracted and translated from largely unstructured patient records, diagnostic exams, pathology reports and discharge summaries by trained medical coders. According to Strong et al. [6], the subjective generation of these values when interpreting the diagnoses can lead to data quality problems, namely loss of objectivity and credibility of the data. In the context of APR-DRG classification, this usually happens when coding errors occurs. At this point, it is necessary to recall that the Portuguese government pays hospitals according to an established list of prices that are linked to each APR-DRG. Therefore, failures in the data representing codes of diseases and procedures will undoubtedly have an economic impact to hospitals.

There are several data quality problems related to medical coding errors that could negatively impact hospital funding. One the most common examples is denominated in the literature as upcoding, which is the practice of miscoding patient data to receive higher reimbursements [7]. Upcoding occurs when coders purposely choose more complex codes than the reality in order to classify patients into higher-complexity APR-DRGs, which in turn will result in more money to the hospitals [8]. That can happen when a hospital coder tries to explore the medical records to extract the most lucrative codes, including changes between the principal and secondary diagnosis, look for reimbursable conditions and exaggerate the choice of codes without supportive evidence in the patient's record, such as adding more diagnoses [9].

Consequently, it is paramount to watch the levels of quality of the APR-DRG data. In this paper, we aimed at assessing the credibility of administrative data regarding coding issues that can potentially impact APR-DRG classification, with emphasis on upcoding. As a pilot for the methodology, we limited our study to inpatient episodes due pneumonia. We assessed the quality of data concerning pneumonia-related diagnoses, which are key codes for properly grouping hospital cases into APR-DRGs in respiratory diseases.

The main contribution of this paper is to describe how to assess the credibility of the dataset used for APR-DRG classification, as we were particularly interested in the degree to which a set of attributes representing diagnosis codes are believable by users, namely hospital providers and managers, regarded as true and how much they represent the reality. In our context, we can attribute low-credibility data when possible upcoding cases are flagged.

2 Methods

2.1 Data Sources

As previously said, data used for this study was extracted from Portugal's National DRG database, which is a nationwide inpatient database containing coded clinical data provided by all public hospitals from the National Health System (NHS) in mainland Portugal. We analyzed data from 2015, which was the last year in which Portugal used the ICD-9-CM to code all episodes. In 2016 onwards, episodes were either coded in ICD-9-CM or in the newest tenth revision, so we opted to avoid further bias related to the transition of ICD versions. We selected all cases with a principal or secondary diagnosis from the ICD-9-CM codes comprised in the interval 480–488, which corresponds to Pneumonia and Influenza diagnoses [10]. All variables required for APR-DRG grouping were collected, namely principal diagnosis, up to 30 secondary diagnoses, up to 30 inpatient procedures, discharge status, sex and age.

Since inpatient data used in this study was completely anonymized and only contained the discharge year, diagnosis and procedure codes, sex, age, discharge status and an arbitrary episode identification number, there was no need for ethical approval.

2.2 Developing the Mechanisms to Measure the Credibility of a Record

To identify abnormal frequencies of APR-DRGs across hospitals, Chi-square test with Bonferroni correction for multiple comparisons was firstly employed. For each discrepant APR-DRG, the same statistical test was employed to compare hospitals regarding the frequency of pneumonia-related diagnosis codes among cases grouped into the discrepant APR-DRGs. Our hypothesis was that patients with the same APR-DRG should not significantly differ in the frequencies of these codes as they present the same hospitalization causes. Following this analysis, all pneumonia codes that accounted for a significantly higher-than-expected frequency in at least one hospital were targeted, as they might lead to suspicious upcoding cases. In order to minimize the bias introduced by differences in the complexity of the patients treated by each hospital, we restricted our analyses to six hospitals with similar capacity based on a standard categorization defined by the Portuguese NHS [11].

To investigate whether the targeted codes could be associated with an upcoding case, we employed Support Vector Machine (SVM) [12] to simulate APR-DRG classification. The main advantage of using SVM is that it can overcome high dimensionality issues [13], which is the case of the APR-DRG classification problem.

We built two SVM-based classification models: (1) one for predicting any APR-DRG related to a disease or disorder of the respiratory system, based upon 17 different APR-DRGs defined in version 31 [5]; and (2) another one to determine the Severity of Illness (SOI) level, which is a score to be added to the APR-DRG, always ranging from 1 to 4 (1 – minor; 2 – moderate; 3 – major; 4 – extreme). The full APR-DRG classification includes both, the APR-DRG itself and the SOI level. We trained the SVM models on two thirds of the inpatient data for the period 2011–2015 and tested their performance on the remaining third. A total of 487,156 cases were used for training and testing the SVM models. As evaluation metrics of the goodness of the model, we considered precision, recall and the percentage of correctly classified cases. We further tested the models on data from the year of 2016 in order to add critical validation to the models and assess their capacity of generalization.

Using the constructed SVM models, we performed a sensitivity analysis to discover the individual role of each discrepant pneumonia code on APR-DRG classification by removing the code from the original dataset and assessing APR-DRG changes. If the exclusion of the code alone moves the episode to a lower intensity APR-DRG, then the episode is labeled as a suspicious case of upcoding. Finally, we estimated the levels of credibility. We define the measure "level of credibility" as the difference of the percentage of possible cases of pneumonia-related upcoding in a given facility from the total number of inpatient episodes in that same facility.

Data processing, training and testing phases of SVM were performed using Java code in combination with a Weka open source library for Java [14], version 3.8.0.

3 Results

Significantly different frequencies of APR-DRG 137 (Major respiratory infections and inflammations) and APR-DRG 139 (Other pneumonia) were found across hospitals. The diagnosis codes that presented a significantly higher-than-expected frequency in at least one hospital were: 482.42 - Methicillin resistant pneumonia due to Staphylococcus aureus (Hospital F), 480.9 - Viral pneumonia, unspecified (Hospital E), 481 - Pneumococcal pneumonia (Hospitals A, C and E), 482.9 - Bacterial pneumonia, unspecified (Hospital A), 485 - Bronchopneumonia, organism unspecified (Hospital B) and 486 - Pneumonia, organism unspecified (Hospitals B, E and F).

Regarding the performance of the SVM-based models, considering the first level (APR-DRG without the SOI level), weighted recall and precision were both 0.994 and the percentage of correctly classified cases were 99.4%. Considering the SOI determination, overall weighted recall and precision were both 0.893, with a percentage of correctly classified cases of 89.3%. When tested in data from 2016 (92475 episodes), we verified that the SVM presented a high capacity of generalization, with a percentage of correctly classified cases of 88.4%.

Table 1 summarizes the sensitivity analysis results by indicating how many episodes were driven to a given APR-DRG by each targeted code. For instance, from 7059 episodes it occurred, code 486 alone was responsible for allocating 6826 episodes to APR-DRG 139 and 1 episode to APR-DRG 137 when coded as principal diagnosis, whereas it accounted for placing 185 episodes (out of 540) once it was coded as secondary diagnosis. In 222 episodes (out of 760), five codes were alone responsible for allocating the episodes into APR-DRG 137 when they are reported as secondary diagnosis rather than principal diagnosis.

Table 1. Individual effects of each discrepant pneumonia-related diagnosis codes on APR-DRG classification

	APR-DRG 137	APR-DRG 139	Total episodes
Principal diagnosis			
482.42	0	0	102
480.9	0	53	54
481	1	483	509
482.9	1	466	488
485	0	449	455
486	1	6926	7059
Total	3	8377	8667
Secondary diagnosis			
482.42	3	0	64
480.9	0	0	5
481	5	0	34
482.9	26	0	78
485	3	0	39
486	185	0	540
Total	222	0	760

Table 2 below shows, for each hospital, the number of hospitalizations flagged as suspicious cases of upcoding and the respective credibility levels. Hospital D was the only one that did not present an abnormal frequency of a pneumonia code. The occurrence of upcoding related to pneumonia was proportionally small, not reaching 1% of the cases in any of the evaluated hospitals. Moreover, the credibility levels across the hospital databases were very high, with values higher than 99%.

Table 2. Levels of credibility concerning upcoding cases in pneumonia

Hospital	Number of suspicious cases of upcoding	Total number of episodes	Credibility level
Hospital A	7 (0.1% of the total)	3884	99.8
Hospital B	51 (0.94% of the total)	5408	99.1
Hospital C	1 (0.05% of the total)	2205	99.9
Hospital E	21 (0.55% of the total)	3846	99.5
Hospital F	49 (0.73% of the total)	6676	99.3

4 Discussion

The credibility of coded clinical data in administrative databases is a critical issue in the context of health care funding, research, decision making and quality of care assessment. The emphasis of this article was to measure the credibility of data concerning upcoding of pneumonia, a condition that already has found to be manipulated in hospital datasets to alter the complexity of hospitalizations in order to increase reimbursements [15–17].

From the clinical point of view, a total of five out of six discrepant codes presented similar effects on APR-DRG grouping as they drove the classification into the APR-DRG 139 as principal diagnosis. The exclusion of these five pneumonia codes would shift these episodes to APR-DRG 137 in nearly all episodes they occurred (8377 out of 8667 episodes, see Table 1), which is an APR-DRG with a higher weight and reimbursement rates [18]. Moreover, in some cases (222 out of 760 episodes, see Table 1), switching these conditions from principal to secondary diagnosis alone would result in more financial compensation to hospitals, as it could prevent episodes from being assigned to APR-DRG 139 and move them to APR-DRG 137 instead. These cases should be watched more closely as they could be an indicator of upcoding practices.

The number of cases flagged as upcoding by our method was proportionally small and the credibility of the data concerning upcoding of pneumonia was very high. The magnitude of upcoding observed in our findings appear to be in line with a systematic literature search conducted by Lüngen and Lauterbach [19], who estimated that upcoding was related with up to 1% of the inpatient care payments in Germany [19]. However, this value is quite lower than the rates identified in a 1995–1996 coding audit in Australia, which revealed that an estimated of 5.2% of the medical records were upcoded [9]. In the United States, it was found that one-third and one-half of the case-mix increase occurred due to upcoding in the periods 1986–87 and 1987–88, respectively [20, 21]. In Portugal, Barros and Braun analyzed the same Portuguese DRG database used in this study and found that upcoding has been occurring in public hospitals to increase their budgets, but the impact was quantitatively small [17].

As a limitation of our study, we mention that flagging possible upcoding cases was based upon results obtained with the direct application of the SVM algorithm. Therefore, existing errors or shortcomings associated with the SVM models might have influenced or been replicated in our results. Furthermore, we only evaluated credibility related to coding the six conditions in which at least one hospital presented a significantly higher-than-expected frequency of cases, and not consider possible coding issues related to other diagnoses or procedures.

5 Conclusion and Future Work

We described and applied a method for monitoring possible upcoding cases related to pneumonia diagnoses. Overall credibility levels of clinical were high and only a few proportions of suspicious cases were flagged by our method. Hospitals significantly differed on reporting six pneumonia conditions that drove the classification to APR-DRG 139 when coded as principal diagnosis, though the episode would move to a higher paying APR-DRG (APR-DRG 137) once these codes are reported as secondary diagnosis. We employed a generic and reproducible method that can be useful for discovering relevant APR-DRG relations and thus to filter cases for audit planning. Future works include the refinement of the machine learning models, including testing different algorithms and approaches, the extension of the proposed methodology to measure other data quality dimensions and other disease domains, automate some part of the process and establish a relationship between the levels of credibility and the amount of reimbursement affected by that.

Acknowledgements. The authors would like to thank the Central Authority for Health Services, I.P. (ACSS) for providing access to the data. We would also like to thank to project GEMA: Generation and Evaluation of Models for Data Quality (Ref.: SBPLY/17/180501/000293) and the Master Programme in Medical Informatics of the Faculties of Medicine and Sciences of the University of Porto for financial support.

References

1. Aiello, F.A., Roddy, S.P.: Inpatient coding and the diagnosis-related groups. J. Vasc. Surg. **66**(5), 1621–1623 (2017)
2. Mathauer, I., Wittenbecher, F.: Hospital payment systems based on diagnosis-related groups: experiences in low- and middle-income countries. Bull. World Health Organ. **91**(10), 746–756 (2013)
3. Cheng, P., Gilchrist, A., Robinson, K.M., Paul, L.: The risk and consequences of clinical miscoding due to inadequate medical documentation: a case study of the impact on health services funding. Health Inf. Manag. J. **38**, 35–46 (2009)
4. Agrupador de GDH All Patient Refined DRG. http://www2.acss.min-saude.pt/Portals/0/CN22.pdf. Accessed 22 May 2019
5. All Patient Refined Diagnosis Related Groups Methodology Overview 3M Health Information Systems. https://www.hcup-us.ahrq.gov/db/nation/nis/grp031_aprdrg_meth_ovrview.pdf. Accessed 22 May 2019
6. Strong, D.M., Lee, Y.W., Wang, R.Y., Strong, D., Lee, Y.W., Wang, R.: 10 potholes in the road to information quality. IEEE Comput. **30**, 38–46 (1997)
7. Dafny, L.S.: How do hospitals respond to price changes. Am. Econ. Rev. **95**, 1525–1547 (2005)
8. Silverman, E., Skinner, J.: Medicare upcoding and hospital ownership. J Health Econ. **23**, 369–389 (2004)
9. Pongpirul, K., Robinson, C.: Hospital manipulations in the DRG system: 755 a systematic scoping review. Asian Biomed. **7**, 301–310 (2013)
10. International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM). https://www.cdc.gov/nchs/icd/icd9cm.htm. Accessed 22 May 2019
11. Administração Central do Sistema de Saúde. Grupos e Instituições. http://benchmarking.acss.min-saude.pt/BH_Enquadramento/GrupoInstituicoes. Accessed 22 May 2019
12. Chu, A., et al.: A decision support system to facilitate management of patients with acute gastrointestinal bleeding. Artif. Intell. Med. **42**, 247–259 (2008)
13. Verplancke, T., et al.: Support vector machine versus logistic regression modeling for prediction of hospital mortality in critically ill patients with haematological malignancies. BMC Med. Inform. Decis. Mak. **8**, 56 (2008)

14. University of Waikato Weka 3: Data Mining Software in Java. https://www.cs.waikato.ac.nz/ml/weka/index.html. Accessed 28 June 2019
15. Sjoding, M.W., Iwashyna, T.J., Dimick, J.B., Cooke, C.R.: Gaming hospital-level pneumonia 30-day mortality and readmission measures by legitimate changes to diagnostic coding. Crit. Care Med. **43**(5), 989–995 (2015)
16. Hebert, P.L., McBean, A.M., Kane, R.L.: Explaining trends in hospitalizations for pneumonia and influenza in the elderly. Med Care Res Rev. **62**(5), 560–582 (2005)
17. Barros, P.P., Braun, G.: Upcoding in a national health service: the evidence from Portugal. Health Econ. **26**, 600–618 (2017)
18. Diário 777 da República. Diário da República, Portaria No. 207/2017 778 de 11 de julho de 2017. http://www.acss.min-saude.pt/wp-content/uploads/2016/12/Portaria_207_2017-1.pdf. Accessed 27 June 2016
19. Lungen, M., Lauterbach, K.W.: Upcoding—a risk for the use of diagnosis-related groups. Dtsch. Med. Wochenschr. **125**, 852–856 (2000)
20. Carter, G.M., Newhouse, J.P., Relles, D.A.: How much change in the case mix index is DRG creep. J. Health Econ. **9**, 411–428 (1990)
21. Carter, G.M., Newhouse, J.P., Relles, D.A.: Has DRG Creep Crept Up? Decomposing the Case Mix Index Change Between 1987 and 1988. RAND Corporation, Santa Monica (1991)