Strategies for Improving the Data Quality in National Hospital Discharge Data System: a Delphi Study

Farahnaz Sadoughi, Zahra Mahmoudzadeh-Sagheb, Maryam Ahmadi
Department of Health Information Management, School of Health Management and Information Sciences, Iran University of Medical Sciences, Tehran, Iran

Corresponding author: Zahra Mahmoudzadeh-Sagheb. Address: Iran University of Medical Sciences, No. 6. Yasemi St., Near Niayesh Highway, Valey-e-Asr Ave., Tehran, Iran. Phone: +982188794300. Mob: +989382455690. Fax: +982188883334. E-mail: zm_222@yahoo.com f-sadoughi@tums.ac.ir.

1. INTRODUCTION

A hospital discharge data system is an informational, planning, budgeting, epidemiologic, and quality control tool in modern healthcare. By using this system and provision of data from all the hospitals, a comprehensive image of the whole population as well as the hospital services used in the region will be provided (1).

Hospital discharge data are in fact the most accessible comparative data resources for investigating all the hospitalized patients (2). Such data enable categorization of the hospitals regarding the comparison of expenditures and facilitate assessing the effect of market competition on the hospital cost growth, planning for better resources allocation, and estimation of the economic burden of diseases (3). These data also play a major role in producing the health indicators (4), community health assessment, disease surveillance, strategic planning, service quality control, research, policy making, and preventing injuries (3).

Of course, hospital discharge data quality affects the usefulness of these data and is one of the prerequisites for effective utilization of the data. Thus, the present study aimed to identify the necessary actions for improving the data quality in the national hospital discharge data system and present a model for Iran based on the experiences of England, Canada, and New Zealand.

Methods: In doing so, the measures performed in these countries were investigated. The related data were organized in six categories of standards and procedures, training and coordination with the users, assurance from the capability of the system’s software, data modification, data quality control, and documentation and reporting the data quality. According to the gathered data, the primary model was designed. Then, the model was assessed using a two-round Delphi technique by 33 and 31 experts, respectively.

Conclusion: According to the findings, a model was presented in order to improve the data quality of Iran’s national hospital discharge data system.

Key words: Data quality, Health information system, Hospital discharge data

2. METHODS

The present study was conducted using descriptive methods and the Delphi technique in 2011-2012. The measures performed in England, Canada, New Zealand, and Iran for improving the data quality in the national hospital discharge data system were investigated. The above-men-
tioned 3 countries were selected due to their long history in establishing the systems, expanding the hospital discharge data system to the national level, appropriate investment for development and design of the systems, and focusing on data quality (10, 11, 12).

The study data were collected by studying the documents, articles, books, and journals either in the library or published in the websites of the health organizations of the selected countries as well as consultation with the experts in the field of hospital discharge data systems of the studied countries via email. The data related to Iran were collected from the publications and instructions issued by the Ministry of Health and the provincial offices of hospital discharge data systems. In this regard, interviews were also conducted with officials and experts.

To design the primary model, the data gathered from the selected countries were organized in 6 categories of standards and procedures, training and coordination with the users, assurance from the capability of the system's software, data modification, data quality control, and documentation and reporting the data quality. These categories were emerged inductively as the literature review was conducted. The gathered data were compared regarding the 6 above mentioned categories. Accordingly, similar items were eliminated and dissimilar ones were included in the primary model. The primary proposed model was designed by cooperation of seven experts, including one PhD of Health Information Management, two PhDs of Medical Informatics, one PhD of Computer Sciences, one physician, and two experts of Medical Records.

Then, the Delphi technique was used to determine the validity of the proposed model. In doing so, a questionnaire was prepared based on the primary model. The questionnaire was sent to 40 experts in health information systems and health information management through E-mail and 33 completed questionnaires were returned (response rate: 82.5%). These experts were the professors of Iran's universities of medical sciences or the authorities of provincial hospital discharge data systems of the country. The participants in the Delphi technique had to score the components of the proposed model using the 5-point Likert scale (1=completely disagree, 5=completely agree) and present their suggestions, as well.

In each round of the Delphi technique, the items with median scores ≥ 4 could remain in the model. According to the results of the first round of Delphi, another questionnaire was prepared for the second round. Then, the questionnaire was sent to the 33 experts who had taken part in the first round of Delphi through E-mail and 31 questionnaires were returned (response rate: 93.9%). After the two rounds, agreement was achieved regarding the components of the proposed model.

### 3. RESULTS

The demographic characteristics of the experts who participated in Delphi technique are shown in Table 1. A total of 33 experts took part in the first round of the Delphi technique and 31 ones participated in the second round. The majority of the experts were faculty members or university professors (72.7 % and 74.2% in rounds one and two, respectively) and the others were the authorities of provincial hospital discharge data systems.

The items of the validated model, including 67 strategies, are presented in 6 categories of standards and procedures, training and coordination with the users, assurance from the capability of the system's software, data modification, data quality control, and documentation and reporting the data quality. After all, 56 strategies based on the experiences of the studied countries and 11 items according to Delphi experts' suggestions were added to the model. Overall, only 5 items of the proposed model have been used in Iran’s current system (Tables 2 to 7).

### 4. DISCUSSION

The results of this research showed that the studied countries (Canada, England, and New Zealand) have undertaken a variety of activities and measures related to the data quality of their hospital discharge data systems. According to the findings, very limited activities have been done in Iran in this regard. Even the items that have been emphasized in all the 3 selected countries, such as utilization of minimum data set at the national level, existence of a comprehensive guide for abstracting and data collection, existence of data
quality control procedures, have been neglected in Iran’s systems. Therefore, this system needs to be reviewed and developed.

Besides, the results indicated that the study experts added 11 new items, such as motivating the staff, providing conditions for preventing the staff from being transferred or leaving the system, and employing specialist workforce in health information management, to the model. This can result in a more complete model.

The most important category of the model is the category of standards and procedures and its strategies should be considered prior to the other presented strategies. However, all the items of models are essential for improving the data quality and their implementation is recommended.

4.1 Standards and procedures

Zadka et al. showed that in order for the national hospital discharge databases to be comparable, existence of sufficient descriptions about these databases which encompass all the comparability disturbing issues and identify the differences is necessary (13). Furthermore, to gather the data in a standard manner, the necessary data should be obtained based on the clear definitions and a standard guideline should also be designed regarding data collection (14, 15, 16). The importance of standardization of the methods has been emphasized by Buffum, as well (17). In fact, standardization of the methods lead to better coordination as well as more desirable outcomes; of course, guidelines should be periodically reviewed. In this regard, Richards et al. emphasized revision of the discharge abstract manual documentation and considered it as a mechanism for supporting the data quality improvement (18). Overall, using a minimum data set in the hospital discharge data system plays a critical role in definition of data elements and creating coordination in data collection, reporting, and usage. In fact, development and utilization of a minimum data set also provides the ground for data quality improvement.

According to Gray and Clement, linkage capability is a basic feature of data sets and is necessary for responding to a great number of research questions. In addition, by predicting the organizational or geographical identifiers and linking to other related data sets, the necessary information can be provided (19). Schoenman et al. considered the capability of linkage to other databases as a criterion for evaluating the hospital discharge databases (3). Thus, the authorities of hospital discharge data systems should establish appropriate linkages.

4.2 Training and coordination with the users

Buffum states that training the data collectors leads to the databases’ readiness for being used in healthcare decisions (17). In the same line, Richards et al. believe that holding educational meetings and conferences is effective in improving the hospital discharge data quality (18). In this regard, holding training courses on coding the diseases is of utmost importance. As Lawrence et al. have mentioned, since the usefulness of hospital discharge data is affected by the quality of diseases coding, presenting a consistent, clear guide about the necessities and expectations to the hospitals and coders is highly important (4). Furthermore, according to Arts et al., a major part of error prevention is related to selection as well as education of sufficient, motivated personnel. They also believe that in order for continuity in error prevention in databases, the staff of the central coordinating center as well as those of the local sites should be motivated (20). Similar to other studies conducted on the issue, the findings of the present study showed that training the users can be highly effective in improving the data quality. Therefore, in addition to holding the common training courses, the authorities of this system have to perform educational needs assessment and hold special training courses for the staff of the hospital discharge data system in both the central office and the local cooperating sites.

4.3 Assurance from the capability of the system’s software

Friedman and Gustafson have emphasized that the utilized software in healthcare organizations should facilitate the users’ quick and timely access to their required data; in a way that no error be there when responding to the information requests (21). After determination of the users’ information requirements and the expected reports, the authorities of hospital discharge data system should reflect the standards as well as the necessary information and consider it as a mechanism for evaluating the hospital discharge databases (3).
Data cleaning refers to the elimination of problems regarding the missing values, incorrect or out of domain measures, and incompatibilities (23). In fact, data anomalies can be revealed by automated data cleaning at data entry (20). In this way, utilizing the parameters of the software program on number values is a method for reducing the errors and in case a number is entered incorrectly, it is revealed by an alarm (17). Of course, all the errors cannot be revealed through automated examination and it is not possible to discover the data errors which exist in the predefined domain. Thus, in addition to automated investigation of the data, a visual examination is recommended, as well (24). For instance, discharge diagnoses can be matched with the medical record in order to manually investigate the software producers. The documents related to the description of the validation rules and the intended features for the files sent to the database should be provided for the software producers, as well (22). Data validation rules are in fact the necessary logical investigations of the data entered to the database in comparison to the predefined rules. These investigations focus on the domain of the values or logical consistency by taking other fields of the patient’s data into account (23). Thus, the authorities should consider development of data validation document as a priority; so that the software can be prepared and the data can be cleaned. Primary and periodical evaluation of the software and applying the necessary modifications lead to data quality improvement, as well.

### 4.4. Data modification

Data cleaning refers to the elimination of problems regarding the

| Category | Strategies | Source |
|----------|------------|--------|
| 1.1 | Automated data cleaning | * |
| 1.2 | Manual data cleaning | * |
| 4.1 | Modifying the logic of the software | * |
| 4.2 | Calculating the errors | * |
| 4.3 | Modifying the data | * |
| 4.4 | Analyzing the data | * |
| 4.5 | Modifying the data | * |
| 4.6 | Analyzing the data | * |

#### Table 5. Strategies related to category of data modification

4.5. Data quality control

Studies have shown that to ensure the usefulness of the hospital discharge data system for achieving scientific goals, it is necessary to evaluate the accuracy and coverage of the related databases (25). Data quality can be controlled by visiting the healthcare institutes and performing data audits in which a sample of the central database data is compared to the source data (14, 26). Moreover, assessment of data collection methods for reducing the probability of bias and error, paying attention to the existing errors, and management of the missing data are other measures which lead to data quality improvement (17).

Furthermore, to validate the data, it is necessary to investigate the monthly activity of healthcare centers in sending data to the national database and compare it to the previous months’ activities. According to Buffum, the measures which have highly increased or decreased compared to the previous ones can be errors. He also believes that comparison of the findings with other databases of the same region or similar organizations can validate the data, as well (17). According to World Health Organization (WHO), if the healthcare ministries or the government of health aim to take serious actions about the data quality, they have to develop a plan for improving the quality of the data as well as the information resulting from the data (27). Consequently, the authorities have to approve and implement a plan for improving the quality of these data, determine an accurate guideline about investigation and controlling the data quality, and monitor its execution.

### 4.6. Documentation and reporting the data quality

According to WHO, documenta-
tion of data quality and publication of the related documents for the users play a critical role in improving the data quality (27). National great databases usually publish documents accurately regarding the data and present explanations about the data source, data collection methods, and reliability and validity of the instruments in these documents (17). By documentation of the data quality, the users will be provided with the opportunity to evaluate the data. According to Schoenman et al., it is hoped that by using this model, the data quality of the national hospital discharge data system will be improved and the system’s objectives will be achieved.

Acknowledgements: This study is a part of a PhD dissertation financially supported by Iran University of Medical Sciences (Grant no: IUMS/SHIMS/7/656), Research Improvement Center of Shiraz University of Medical Sciences, Shiraz, Iran and Ms. A. Keivanshekouh are appreciated for improving the use of English in the manuscript.

5. CONCLUSION

In this research, a model was presented in six categories for improving the data quality of the national hospital discharge data system. The proposed model was a combination of the experiences of the studied countries completed by the suggestions of the experts taking part in the Delphi technique. The model provides some strategies for guiding the authorities of national hospital discharge data system to develop and complete the system regarding the data quality. Overall, establishing a national committee for developing the policies and standards and organizing a team for regularly auditing the data quality can provide the ground for execution of the proposed model. It is hoped that by using this model, the data quality of the national hospital discharge data system will be improved and the system’s objectives will be achieved.

REFERENCES

1. Pulkhinsky TH, Varazkova EA. The new public health, an introduction for the 21st century. 2nd Edition, Sandiego, Academic Press. 2008.
2. Schwartz RM. et al. Administrative data for quality improvement. Pediatrics. 1999; 103(2): 291-202.
3. Schoenman JA. et al. Understanding and enhancing the value of hospital discharge data. Medical Care Research and Review. 2007; 64(4): 440-468.
4. Lawrence BA. et al. Issues in using state hospital discharge data in injury control research and surveillance. Accident analysis and prevention. 2007; 39(4): 319-325.
5. Vestberg K. et al. Data quality of administratively collected hospital discharge data for liver cirrhosis epidemiology. Journal of Medical Systems. 1997; 21(5): 11-20.
6. Muri S. et al. Systemic factors of errors in the case identification process of the national routine health information system: A case study of Modified Field Health Services Information System in the Philippines. BMC Health Services Research. 2011; 11: 271. doi: 10.1186/1472-6963-11-271.
7. Ahmad M, Hoseini F, Barabadi M. Assessment of hospital information systems compliance with the needs of medical records users. Health Management. 2008; 11(12): 25-32 [In Persian].
8. Khaledi N, Ahmadi M, Ayatollahi H. A study on the attitudes of users about application of Admission Discharge software (ADS-Q). Hakim. 2008; 11(5): 47-53 [In Persian].
9. Ministry of Health & Medical Education, Statistics and Information Technology Office. Inpatients Diseases Status in Iran’s Hospitals. Tehran: Ministry of Health & Medical Education, 2009 [In Persian].
10. What is HES. The NHS Information Centre for health and social care. 2011 [http://www.hesonline.nhs.uk/Ease/serve?con tentServer?siteID=1097&categoryID=426, Accessed 17 May 2012].
11. DAD Abstracting Manual (2010-2011 Edition), Ottawa, Ontario, Canadian Institute for Health Information (CIHI), 2010.
12. National minimum dataset (Canada, Statistical events). New Zealand Ministry of Health, 2010. [http://www.health.govt.nz/nz-health-statistics/national-collections-and-surveys/collections/national-minimum-dataset-hospital-events]. Accessed 23 July 2011.
13. Zadk P et al. Hospital Discharge National Databases Pilot questionnaire design testing and results. 1999 [http://www.ncbi.nlm.nih.gov/pubmed/9945758, Accessed 17 May 2012].
14. Whitney CW, Lind BK, Wahl PW. Quality assurance and quality control in longitudinal studies. Epidemiologic Reviews. 1998, 20(1): 71-80.
15. Gasman J. et al. Data quality assurance, monitoring, and reporting. Controlled Clinical Trials. 1995; 16 (2). Suppl. 104:1365.
16. Clive RE. et al. A national quality improvement effort: Cancer registry data. Journal of Surgical Oncology. 1993, 53(5): 155-161.
17. Buffum MD. Databases: Accuracy and readiness for use. Nursing and Health Science. 2011, 2(3): 123-130.
18. Richards J, Brown A, Homan C. The data quality study of the Canadian discharge abstract database. Proceedings of Statistics Canada Symposium, 2001. [http://www5.statcan.gc.ca/bsolc/olc-cel/olc-cel?catno=11-522- X2001012&lang=eng, Accessed 23 November 2012].
19. Gray BH, Clement P. Databases for research on nonprofit health care organizations, opportunities and limitations. American Behavioral Scientist. 2002; 45(10): 1550-1561.
20. Arts DG, de Keizer NF, Scheffer GJ. Defining and improving data quality in medical registries: a literature review, case study, and generic framework. Journal of the American Medical Informatics Association. 1995; 5(6): 600-611.
21. Friedman RB, Gustafson DH. Computers in clinical medicine. Computers and Biomedical Research. 1995, 14(6): 209-204.
22. Data submission requirements, Discharge Abstract Database (DAD), Version 1.0 (2009-2010). Ottawa, Ontario, Canadian Institute for Health Information (CIHI), 2009.
23. Gliklich RE, Dreyer NA. eds. Registries for Evaluating Patient Outcomes: A User’s Guide, and ed. Agency for Healthcare Research and Quality (AHRQ), 2010.
24. Horbar JD, Leahy KA. An assessment of the case identification process of the national minimum dataset-hospital events. New Zealand Ministry of Health, 2010. [http://www.health.govt.nz/nz-health-statistics/national-collections-and-surveys/collections/national-minimum-dataset-hospital-events]. Accessed 23 July 2011.

Table 7. Strategies related to category of documentation and reporting the data quality

| Category | Strategies | Source |
|----------|------------|--------|
| 61 | Documentation of causes and disseminating documents for the users | No |
| 62 | Documenting data quality reports to the data provider | No |
| 63 | Availability of data quality documentation and reporting the data coverage | No |
| 64 | Providing descriptive notes about data quality when publishing the data | No |
| 65 | Presenting data quality reports to the data provider | No |
| 66 | Presenting national reports when checking the data quality | No |
| 67 | Existence of data quality database and the possibility of computing the nursing data quality with the national and regional average through the website | No |
| 68 | Documentation and reporting the data quality | No |
| 69 | Reporting the data quality report | No |
| 70 | Documentation of the changes applied to the database over time | No |
| 71 | Clear statement of the relevant population in all publications | No |
| 72 | Reporting the validity of the key fields based on the period | No |

ACTA INFORM MED. 2013 DEC; 21(4): 261-265 / ORIGINAL PAPER