

**ASSESSMENT OF DATA QUALITY AND RECOMMENDATIONS TO IMPROVE QUALITY OF SHIRAZ UNIVERSITY OF MEDICAL SCIENCES HEMODIALYSIS DATABASE**Neda Firouraghi<sup>1</sup>, Shahrokh Ezzatzadegan Jahromi<sup>2</sup>, Ashkan Sami<sup>3</sup>, Roxana Sharifian\*<sup>4</sup>

1: MSc of Medical Informatics, Shiraz University of Medical Sciences, Shiraz, Iran

2: Assistant Professor of Medicine, Shiraz Nephrology Urology Research Center, Nemazee Hospital, Shiraz University of Medical Sciences, Shiraz, Iran

3: Assistant Professor of Computer Sciences, Shiraz University, Shiraz, Iran

4: Associate Professor, Health Human Resources Research Center, Faculty of Management and Medical Information Sciences, Shiraz University of Medical Sciences, Shiraz, Iran

Correspondence:

Tel: +98 917 315 4688, E-mail: [Sharifianr@sums.ac.ir](mailto:Sharifianr@sums.ac.ir)**TYPE OF ARTICLE:** CONFERENCE ABSTRACT**ABSTRACT**

**Introduction:** Clinical data contain abnormalities, so quality assessment and reporting of data errors are necessary. Data quality analysis consists of developing strategies, making recommendations to avoid future errors, and improving the quality of data entry by identifying error types and their causes. Therefore, this approach can be extremely useful to improve the quality of the databases. The aim of this study was to analyze a hemodialysis (HD) database in order to improve the quality of data entry to avoid future errors.

**Methods:** The study was done on Shiraz University of Medical Sciences' HD database in 2015. The database consists of 2367 patients who had at least 12-months follow-up ( $22.34 \pm 11.52$  months) from 2012–2014. Duplicated data were removed; outliers (out of expected and acceptable range values for each HD variable), based on expert opinion and relationship between variables, were detected. After removing the outliers, missing values were handled by mean (for continues variables) and mode (for categorical variables) in 72 variables by using IBM SPSS Statistics 22. Some recommendations were given to improve the data-entry process, according to the results, error types, and their causes.

**Results:** The variables had outliers in the range of 0–9.28%. Seven variables had missing values over 20% and the others had 0–19.73%. The majority of missing values belong to serum alkaline phosphatase, uric acid, high and low density lipoprotein, total iron binding capacity, hepatitis B surface antibody titer, and parathyroid hormone. The variables with displacement (the values of two or more variables were recorded in the wrong attribute) were weight, serum creatinine, blood urea nitrogen, and systolic and diastolic blood pressure. These variables may lead to decreased data quality. According to the results and expert opinions, applying some data entry principles such as defining ranges for values, not permitting duplicated data or empty field, using the relationship between hemodialysis features, developing alert systems about empty or duplicated data, and entering directly HD data or lab results into the database can improve the data quality drastically.

**Conclusion:** Expert opinions in detecting outliers as a complement to statistical methods can play an effective role in the detection of real outliers. For the analysis of HD databases, the relationship between variables due to their effect on the quality, data entry principles, and mentioned variables should be focused to improve the database quality.

**KEYWORDS:** Assessment, Database, Data entry, Hemodialysis, Quality, Recommendation

**Abstracts of First National Congress of Medical Informatics, Mashhad, Iran, February 2017**

© 2017 The Authors. This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.