MEDINFO 2021: One World, One Health – Global Partnership for Digital Innovation
P. Otero et al. (Eds.)

© 2022 International Medical Informatics Association (IMIA) and IOS Press.
This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0).
doi:10.3233/SHTI220232

How Is the Quality of Information in a Cancer Hospital Regarding the Completeness and Consistency? A Case Study of the Hospital Cancer Registry of the Brazilian Hospital Bom Pastor

Erika G. Assis[a], Anna Carolina L. Lima[b], Cristiane Neri Nobre[a]

[a]*Graduate Program in Informatics, Pontifical Catholic University of Minas Gerais, Belo Horizonte, Minas Gerais, Brazil*
[b]*Promove Faculty of Technology, Belo Horizonte, Minas Gerais, Brazil*

Abstract

This study aimed to analyze the measurement of information quality in the Information System of the Brazilian Hospital in terms of data completeness and consistency. The quality of information is a fundamental factor in evaluating the performance of information systems. Its use is a reasonable condition in the management of the health service and attention to the care of patients. The research had a retrospective, descriptive character exploratory with quantitative analysis of the data. The study had some limitations, we observed incomplete information regarding some studied variables; this is because the primary source of information in the cancer registry is the patient's medical record. Therefore, the patient's medical records are relevant and should contain the entire health history, from birth to death. In addition, they support research, the management of health services.

Keywords:

Data Accuracy, Hospital Information Systems, Information Technology.

Introduction

Cancer is a worldwide public health problem, especially in developing countries. The incidence of the disease in the population is expected to correspond to 80% of the more than 32.2 million new cases estimated for 2040 [9].

The information generated from hospital records is essential for the knowledge of the population's health situation. From this information, it is possible to establish subsidies for excellent care to the public, creating specific public policies and formulation of health programs [1] [10]. According to Empel [5], the data quality is fundamental because the data subsidizes the information; that is, the information is the base for creating the knowledge. Thus, this study aimed to analyze the measurement of information quality in the Information System of a Brazilian Hospital in terms of data *completeness* and *consistency*. The quality of information is a fundamental factor in assessing the performance of information systems, and its use is a fundamental condition in the management of the health service and attention to patient care.

Completeness refers to the full description/completion of a standard about its nature; that is, the degree of completion of the analyzed field is measured by the proportion of notifications to completion [3] [9]. The consistency was analyzed by evaluating the size of cases filled with codes that did not exist in the Tumor Registration Form or with incorrect filling formats and the proportion of records with inconsistency between the variables analyzed.

Methods

The research had a retrospective, descriptive exploratory character with quantitative analysis of the data [9]. The investigation consisted of measuring the quality of the Hospital Cancer Registry of the Brazilian Hospital by assessing the completeness and consistency of the data. The cancer registry activities follow the following steps: Identification of Cases, Analysis and Registration of Tumors, and Hospital Registration of Cancer. The registration of cancer is done by registrars who receive technical training from INCA through the technical training course in cancer registration. The tumor registration form is used to collect data from medical records. In this study, the classification used by [3], [6], [9], [11], [12] and [15] was used to assess the degrees of non-completeness and non-consistency of the variables: *excellent* (less than 5%), *good* (5 to 10%), *fair* (10 to 20%), *bad* (20 to 50%) and *very bad* (50% or more).

Results

Medical records of patients from the year of diagnosis until 2011 were analyzed: those who had at least five years of follow-up of the tumor, so they are records until the year 2015. According to the secondary data analysis extracted from the SISRHC, were 9463 cancer records of 9134 patients [2].

The educational level variable presented an abysmal completeness rating with 72% without information. Regarding education, research shows that women with less than eight years of study tend to undergo mammography less frequently and to have a higher prevalence of breast cancer [4] [7] [14].

Relating the survival rate to schooling, studies show that women with higher education have a global survival in five years equivalent to 92%; for women with education up to high school, 84%; 1st degree, 74%; and illiterate, to 56%. Illiterate women are about seven times more likely to die than college-educated women. The increase in the level of education points to a tendency to decrease the risk of death [8]. Regarding occupation, completeness was considered bad, with 49% of the records without information.

The primary tumor location presented excellent completeness, with only 2% of the records with unknown primary location (ICD 80) and without inconsistencies.

Regarding the TNM classification that informs the graduations for the extension of the primary tumor (T), the spread of the disease to regional lymph nodes (N), and the presence of distant metastases (M). The completeness rating was abysmal, with 51% of the records without information. TNM and staging are essential to assess patient care and support studies on survival.

They are important prognostic factors for the evolution of the disease and are also very useful in early detection.

Regarding the primary cause of death, 97.8% of the records are without information showing that they were very poor completeness.

The family history of cancer presented very poor completeness, with 78% of the patients without information. Heredity is also a significant risk factor; for example: if a first-degree relative has the disease, the risk is at least twice as high for the individual to have prostate cancer, and if two or more individuals in the same family are affected, the risk increases by 5 to 11 times [8].

Regarding alcohol consumption, completeness was considered bad, with 49% of patients without information. Concerning smoking, completeness was also bad: 43% of patients do not have information about smoking.

In this study, inadequate completeness was observed for the variables race, occupation, staging, alcohol consumption, smoking. Very bad completeness was noted in the following variables: education, TNM, the primary cause of death, and family history of cancer. It is observed that the factors associated with an increased risk of developing cancer, which is: smoking, alcoholism, and family history of cancer [8] and [13], presented bad or very bad completeness. The variables staging, level of education, and presence of metastases are also related to better survival in patients diagnosed with cancer [14] and presented bad or very bad completeness.

Conclusion

The objective of the work, which was to analyze the measurement of information quality in the Information System of Hospital Cancer of Hospital in terms of data completeness and consistency, was achieved. A lot of incomplete information was observed regarding some studied variables. Therefore, data must be grouped through the adoption of governance principles since the primary purpose of information systems is to guarantee the quality and reliability of the data.

Data quality is one of the main concerns in data learning, for example, since most of these algorithms induce knowledge strictly from the data; therefore, the knowledge acquired is negatively influenced by the lack of data. In this way, this work contributes for future analysis to be carried out in hospital systems to obtain systems with higher quality in terms of the aspects analyzed in this work.

Acknowledgements

This study was financed in part by the Coordination for the Improvement of Higher Education Personnel - Brasil (CAPES) - Finance Code 001. The authors thank the National Council for Scientific and Technological Development of Brazil (CNPq - Conselho Nacional de Desenvolvimento Científico e Tecnológico) and the Foundation for Research Support of the Minas Gerais State (FAPEMIG). The work was developed at the Pontifical Catholic University of Minas Gerais, PUC Minas in the Applied Computational Intelligence laboratory – LICAP.

References

[1] Brasil. Ministério da Saúde. Departamento de Ciência e Tecnologia, Secretaria de Ciência e Tecnologia e Insumos Estratégicos. Integração de informações dos registros de câncer brasileiros. Rev. Saúde Pública, São Paulo, v. 41, n. 5, p. 865-868, Oct. 2007.

[2] Brasil. Ministério da Saúde, Instituto Nacional do Câncer. Registro Hospitalar de Câncer: Dados das Unidades Hospitalares do INCA – 1992 Rio de Janeiro: Pro Onco, 1994. 77p.

[3] Calazans, A.T.S. Information quality: concepts and application. *TransInformação*, Campinas, v. 20, n.1, p.29-45, jan. /abr., 2008.

[4] Felix JD et al. Avaliação da completude das variáveis epidemiológicas do sistema de informações sobre mortalidade em mulheres com óbitos por câncer de mama na região sudeste do Brasil (1998 a 2007). *Ciência e Saúde Coletiva*, Rio de Janeiro, v. 17, n. 4, p. 945-953, 2012.

[5] Globocan 2020: Estimated Cancer Incidence, Mortality and Prevalence Worldwide in 2020. Available from: http://globocan.iarc.fr/Pages/fact_sheets_population.aspx

[6] Gould JB. Vital records for quality improvement. *Pediatrics*, v.103, n.1, Suppl E, p.278-90, 1999. Available at: https://bityli.com/lyWUWi

[7] Hofelmann DA, Anjos JC, Ayala AL. Sobrevida em dez anos e fatores prognósticos em mulheres com câncer de mama em Joinville, Santa Catarina, Brasil. *Ciênc. Saúde coletiva*, Rio de Janeiro, v. 19, n. 6, p. 1813-1824, Jun 2014.

[8] INCA. Estimativa 2020: incidência de câncer no Brasil /INCA– Rio de Janeiro : INCA, 2019. Available at: https://www.inca.gov.br/

[9] Lima CRA et al. Revisão das dimensões de qualidade dos dados e métodos aplicados na avaliação dos sistemas de informação em saúde. *Cad Saude Pública,* v.25, n.10, p.2095-2109, 2009.

[10] Medeiros FGS. O sistema de informação de atenção básica – SIAB no programa de saúde da família – PSF. 2001. 133 f. Dissertação (Mestrado) - Escola da Ciência da Informação, Universidade Federal de Minas Gerais, Minas Gerais, 2001.

[11] Pinto IV et al. Completude e consistência dos dados dos registros hospitalares de câncer no Brasil. *Cad. Saúde Colet.* 2012, Rio de Janeiro, v. 20, n.1, 2012, p.113-20. Available from:http://www.iesc.ufrj.br/cadernos/images/csc/2012_1/artigos/CSC_v20n1_113-120.pdf

[12] Romero DE, Cunha CB. Avaliação da qualidade das variáveis socioeconômicas e demográficas dos óbitos de crianças menores de um ano registrados no Sistema de Informações sobre Mortalidade do Brasil (1996/2001). *Cad. Saúde Pública*, Rio de Janeiro, v. 22, n. 3, p. 673-681. Available at https://bityli.com/UysxTY

[13] Simeão FAP et al. Qualidade de vida em grupos de mulheres acometidas de câncer de mama. *Ciênc. Saúde coletiva,* Rio de Janeiro, v. 18, n. 3, p. 779-788, mar. 2013. Available from: http://www.scielo.br/pdf/csc/v18n3/24.pdf

[14] Stafin I, Caponi LGF, Torres TP, Araujo JN, Guedes VR. Fatores prognósticos no câncer de mama. *HU Revista*, Juiz de Fora, v. 38, n. 2, p. 47-53, abr./jun. 2012.

[15] Wong D, Wub N, Watkinson P. Quantitative metrics for evaluating the phased roll-out of clinical information systems. *International Journal of Medical Informatics*, v.105, p. 130–135, 2017. Available at: https://bityli.com/XMKwnc

Address for correspondence

Érika G de Assis - dudabh@gmail.com

Cristiane N. Nobre - nobre@pucminas.br