National Health Card: Assessment of Reliability of Oncology Encoded Databases

doi: https://doi.org/10.32635/2176-9745.RBC.2019v65n2.337

Cartão Nacional de Saúde: Avaliação da Confiabilidade em Bases de Dados Codificadas da Oncologia

Tarjeta Nacional de Salud: Evaluación de la Conformidad en Bases de Datos de la Oncología Codificadas

Adriana Tavares de Moraes Atty[1]; Jeane Glaucia Tomazelli[2]; Maria Beatriz Kneipp Dias[3]; Caroline Madalena Ribeiro[4]; Arn Migowski[5]; Neilane Bertoni[6]

Abstract

Introduction: The National Health Card (CNS) was created to be an instrument of individual identification of the citizen. In the databases provided by DATASUS, the variable CNS is encoded. **Objective:** To verify if the encoded variable of the CNS in public databases provided by DATASUS can be used as a unique identifier of the patient. **Method:** It was conducted a deterministic linkage between the High Complexity Procedures in Oncology (Apac-onco) database of chemotherapy and radiotherapy, provided by DATASUS, and the databases of an oncology reference hospital located in the State of Rio de Janeiro, considering data from 2010 to 2016. **Results:** In the database of chemotherapy, 2.83% of the CNSs had more than one encoding and, in radiotherapy's, 0.55% similarly. Consequently, the utilization of the CNS encoded exceeded the count of 45.5% of the 77 CID of the chemotherapy database and 20.2% of the 84 CID of radiotherapy. **Conclusion:** In the current format, the CNS encoded does not ensure a univocal identification of SUS patients, making it difficult to follow up the procedures, to estimate the treatment costs, identify barriers to access and plan the health care system organization.
Key words: Confidentiality; Information Systems; Unified Health System; Databases as Topic.

Resumo

Introdução: O Cartão Nacional de Saúde (CNS) foi criado para ser um instrumento de identificação individualizada do cidadão. Nas bases de dados disponibilizadas pelo Departamento de Informática do Sistema Único de Saúde (DATASUS), a variável CNS é codificada. **Objetivo:** Verificar se a variável CNS, codificada nas bases de dados do DATASUS, pode ser utilizada como identificador unívoco do usuário. **Método:** Realizou-se relacionamento determinístico entre a base de Autorizações de Procedimentos de Alta Complexidade em Oncologia (Apac-onco) de quimioterapia e radioterapia do DATASUS e as de um Hospital referência em oncologia no Estado do Rio de Janeiro, entre 2010 a 2016. **Resultados** Na base de dados de quimioterapia, 2,83% dos CNS tinham mais de uma codificação, e, na de radioterapia, 0,55% também apresentavam mais de uma codificação. Consequentemente, a utilização da CNS codificada excedeu a contagem de 45,5% das 77 CID no banco de quimioterapia e 20,2% das 84 CID no de radioterapia. **Conclusão:** No formato atual, o CNS codificado não garante identificação unívoca dos usuários, dificultando o acompanhamento dos procedimentos, a estimativa de custos de tratamento, a identificação de barreiras de acesso e o planejamento da organização da rede de atenção à saúde.
Palavras-chave: Confidencialidade; Sistemas de Informação; Sistema Único de Saúde; Bases de Dados como Assunto.

Resumen

Introducción: La Tarjeta Nacional de Salud (CNS) fue creada para ser un instrumento de identificación individualizada del ciudadano. En las bases de datos proporcionadas por DATASUS, la variable CNS está codificada. **Objetivo:** Verificar si la variable CNS codificada en las bases de datos de DATASUS puede ser utilizada como identificador unívoco del usuario. **Método:** Se realizó una relación determinista entre la base de Autorizaciones de Procedimientos de Alta Complejidad en Oncología (Apac-onco) de quimioterapia y radioterapia, de DATASUS, y las de un Hospital referencia en oncología en el Estado de Río de Janeiro, entre 2010 a 2016. **Resultados:** En la base de datos de quimioterapia, 2,83% de los CNS tenía más de una codificación, y en radioterapia, el 0,55% también tenía más de una codificación. En consecuencia, la utilización de la CNS codificada sobreestimó el conteo del 45,5% de las 77 CID en el banco de quimioterapia y el 20,2% de las 84 CID en la radioterapia. **Conclusión:** En el formato actual, el CNS codificado no garantiza identificación unívoca de los usuarios, dificultando el seguimiento de los procedimientos, la estimación de costos de tratamiento, la identificación de barreras de acceso y la planificación de la organización de la red de atención a la salud.
Palabras clave: Confidencialidad; Sistemas de Información; Sistema Único de Salud; Bases de Datos como Asunto.

[1] Instituto Nacional de Câncer José Alencar Gomes da Silva (INCA). Rio de Janeiro (RJ), Brazil. Orcid iD: https://orcid.org/0000-0003-2271-746X
[2] INCA. Rio de Janeiro (RJ), Brazil. Orcid iD: https://orcid.org/0000-0002-2472-3444
[3] INCA. Rio de Janeiro (RJ), Brazil. Orcid iD: https://orcid.org/0000-0002-5847-9830
[4] INCA. Rio de Janeiro (RJ), Brazil. Orcid iD: https://orcid.org/0000-0003-2690-5791
[5] INCA. Rio de Janeiro (RJ), Brazil. Orcid iD: https://orcid.org/0000-0002-4861-2319
[6] INCA. Rio de Janeiro (RJ), Brasil. Orcid iD: https://orcid.org/0000-0002-2539-9965
Address for Correspondence: Adriana Tavares de Moraes Atty. Rua Marquês de Pombal, 125, 7º andar – Centro. Rio de Janeiro (RJ), Brazil. CEP 22.230-240. E-mail: aatty@inca.gov.br

INTRODUCTION

The National Health Card (CNS) was created to be an instrument of individualized identification of the citizen to follow up its trajectory through the National Health System (SUS), reorienting the model of care to health[1-3]. Its use in the Health Information System (SIS) allows knowing the user's profile, the obstacles of the health services network, SUS coverage, further to the references across the federation entities[2].

The implementation of the CNS has been gradual throughout time. Created by Directive number 1,560 dated 2002[4], in 2004[5], became mandatory for the procedures conducted in wards (Authorizations of High Complexity Procedures - Apac) and hospitals (Authorizations of Hospital Admissions – AIH) that needed previous authorization. The regulation of the law occurred in 2011, being assigned to the National Health System Computing Department (DATASUS) the task of verifying the information signed-in and match them with CNS database to detect duplicities and registration inconsistencies [3].

In order to promote the organization of regionalized and hierarchical services, several inter-operational standards and information on health were regulated within SUS, of complementary health and private systems to integrate health systems, legated or new. Containing unique qualification and identification data of CNS per individuals, it would be possible to relate the information of SUS users in different systems of information to implement one national and longitudinal Health Electronic Register (RES)[6].

The qualification of the CNS base foresees a process of standardization and correction of data, which includes the connection with the database of the Internal Revenue System[7]. Once applied the criteria of quality, it is assigned a master number to the individual and to this number are added the other CNS it may have. This process is called "cleaning of database", is continuous and conducted exclusively by DATASUS.

Information about the attendance provided by SUS are available at DATASUS web address www.datasus.gov.br, with no access restrictions, but protecting the right to privacy and secrecy of the identification of the users through a code of the variable CNS[7,8]. Therefore, the public power ensures the protection of the information, its availability, authenticity and integrality at the same time that promotes the disclosure of information of collective interest produced or guarded by itself [9].

To have the possibility of following up the therapeutic trajectory of the user, evaluate the quality of the assistance and improve the network of attention to health through several free-access SIS would be a conquest for the management of SUS. Therefore, the objective of this study was to verify whether the coded variable related to CNS available in DATASUS databases, could, actually be utilized as unequivocal identifier of users in studies with individualized data.

METHOD

A deterministic link was created between the Apac-onco database available at the Internet by DATASUS and the database of a reference hospital, the 'Instituto Nacional de Câncer José Alencar Gomes da Silva (INCA)", licensed in Oncology of the State of Rio de Janeiro. The data address the chemotherapy and radiotherapy procedures conducted between 2010 and 2016. The link was performed with the variable "number of the authorization" (number of the Apac).

Apac-onco database offered by DATASUS has the variable "Coded National Health Card", which hamper the identification of the user. It were selected solely the initial authorizations of chemotherapy and radiotherapy, Apac type 1, considering only that these represent the first authorization of the patient within the validity period of Apac. It were utilized other variables as: number of the authorization of Apac, coded CNS, gender, code of the International Classification of Diseases and Health Related Problems (ICD-10) and the code of the health facilities registered at the National Registration of Health Facilities (CNES of the facility). The Apac-onco database was downloaded on March 23, 2018.

The database with the production of the Apac-onco obtained from the reference hospital was identified with the number of the SUS card (CNS), name of the patient, gender, date of birth, number of authorization of Apac and CNES of the health facility, in addition to other clinical information as ICD-10 and year of the treatment.

To evaluate the unicity of the codes assigned by DATASUS to CNS, the possible situations were investigated: (a) the occurrence of more than one number of CNS with the same code; (b) occurrence of more than one code for a same number of CNS. It was evaluated as well if, for the same authorization Apac there were differences of the variables related to the principal ICD and to the patient gender across the bases.

Another strategy was to evaluate the national base of DATASUS of chemotherapy and radiotherapy and investigate whether a same coded CNS appeared in registers of different genders and municipalities of the health facility where the treatment was conducted.

All the analyzes were performed with the *software R v.3.5.0*, utilizing the package *tidyverse*[10]. The Institutional

Review Board of INCA approved the study, number CAAE: 77845317.9.0000.5274.

RESULT

After the deterministic linkage, 298,028 Apac registers were obtained corresponding to 79,228 CNS at the hospital base (CNS) and 80,983 at DATASUS base (coded CNS). The registers were grouped according to the therapeutic modalities: chemotherapy, radiotherapy and chemotherapy concomitant to radiotherapy (Table 1).

In the database of radiotherapy, it were identified 106 CNS with more than one code, which represents 0.55% of the cards (106/19.341). In the chemotherapy bank, this proportion was 2.83% (712/25.114). In the combined bank (chemotherapy and radiotherapy), the proportion of cards with more than one code associated was 2.67% (928/34,773). No case was identified where the same code has been utilized for more than one CNS.

Considering that, for the same CNS, different codes in the base of DATASUS were assigned, the utilization of the coded variable CNS to represent one case instead of using the identified register of CNS ended up overestimating the counting of 45.5% of the 77 ICD of the chemotherapy bank and 20.2% of the 84 ICD of the radiotherapy bank. For chemotherapy, the overestimation varied from 0.16% to 5.88% while in radiotherapy, it was from 0.17% to 3.07%, depending of the ICD (Figure 1). The most overestimated ICD in chemotherapy was malignant neoplasm of the rectosigmoid junction (C19) and in radiotherapy, was malignant neoplasm of the vulva (C51).

Among the selected registers by the number of the authorization, the ICD of neoplasm reported was discrepant in two registers, the hospital base and DATASUS base. The variable gender differed in 99 registers, corresponding to 33 CNS.

While evaluating the national data of DATASUS, in what concerns Apac-type 1 alone, the chemotherapy bank presented for the period in study, 1,221,899 distinguished coded CNS, of which 13,837 (1.13%) appeared in more than one municipality for treatment. The national radiotherapy bank had 700,368 coded CNS and 9,790 (1.40%) appeared in health facilities of different municipalities.

Table 1. Total registers in the databases after deterministic linkage of the hospital base identified and DATASUS bank, 2010 to 2016

Banks	Authorizations of High Complexity Procedures	National Health Card	Coded National Health Card
Chemotherapy	124.836	25.114	25.830
Radiotherapy	24.178	19.341	19.447
Chemotherapy + Radiotherapy	149.014	34.773	35.706

Source: Ward Information System - SIA/SUS. INCA Database.

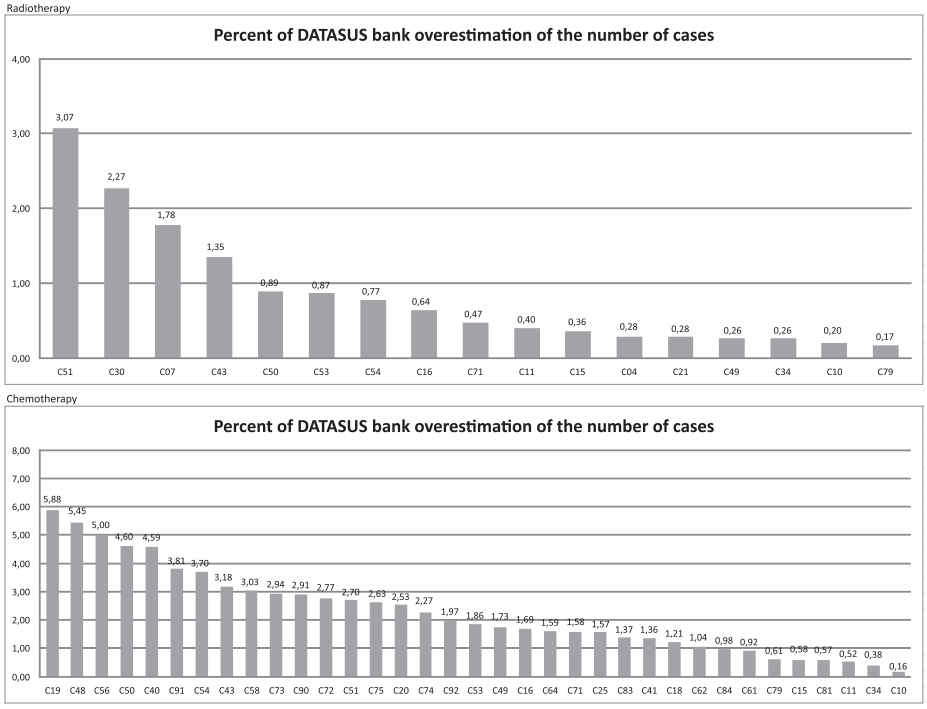

Figure 1. Percent of overestimation of the number of cases when utilized the coded CNS of DATASUS base, 2010 to 2016

The same code associated to different genders registers occurred in 329 cases in the bank of radiotherapy, representing 0.05% of the CNS codes and in the chemotherapy bank, this figure reached 4,258 coded CNS (0.35%).

DISCUSSION

There is a debate about the methods of linkage across health databases, considering that they can be deterministic or probabilistic[11-13]. In the absence of a univocal identifier, most often the probabilistic tracking has considerably superior sensitivity than the deterministic [14]. Several countries, especially in Europe and in the United States of America adopted univocal identifiers in health records, either using specific health numbers or others used for diverse finalities domestically[15]. The existence of a unique identifier allows deterministic linkage of several registries of a same individual more effectively and reliably. Nonetheless, the implantation of these identifiers is a major challenge for low and medium income countries [16]. The process developed in the current article utilized deterministic linkage to attempt to establish a univocal link across the database analyzed, allowing to discuss the reliability of CNS coding as unique identifier of DATASUS databases.

This study identified that, so far, the coded CNS utilized in DATASUS database does not ensure the univocal identification of SUS users as established in the Directives of the Ministry of Health. Even holding a majoritarian correspondence between the information of the bases utilized, issues have been identified. Although the proportion of CNS with more than one code had been under 3%, this type of error can yield important biases while monitoring the assistance granted to the user, utilizing a same information system or combination of information sourced from different SIS.

Several studies [17-21] utilize bases of public access of DATASUS and consider the coded key-variable to define "case" and describe the profile of the assistance. Other studies[22-26] utilize identified databases, but have difficulties of obtaining these bases. Therefore, the availability of the coded databases is an alternative to perform investigations, waiving the cession of the nominally identified base and ensuring the confidentiality of the data.

The utilization of coded bases with the problems identified may generate distorted results since the non-unicity of CNS codes, besides hampering the complete follow up of the individual by the health system, inflates the quantity of cases. It was observed that the most overestimated ICD do not match the most frequent neoplasms in Rio de Janeiro, Brazil, discarding a possible explanation of a bigger number of differences connected to higher occurrence of neoplasm.

Another problem identified was the occurrence of the variable "male gender" in one base and "female" in another, when the variable SUS card corresponded to the coded variable. Similar situation occurred for the coded variable of neoplasm – ICD. It is possible that the difference observed for the variable gender is a consequence of the adjustment of the national base after DATASUS review[3,5]. However, this adjustment does not explain the difference observed in the register of ICD.

The cases where the same coded CNS appeared in health facilities from different municipalities, though possible, do not allow us to affirm it corresponds to the therapeutic trajectory the patient has gone through because the study showed that the same CNS code was used for different individuals.

The main limitation of this study is his scope, because it were used the data of only one hospital. Regardless of this hospital presents the biggest local production in this area and is a reference for the entire State, it is necessary to broaden this investigation to verify whether the results obtained replicate for different hospitals.

CONCLUSION

The coding of CNS in every information system is mandatory to ensure the privacy of the users' data of SUS. However, it is necessary to unify this coding. As such, it will be possible to verify the intervals among the procedures foreseen in the lines of care to estimate the costs of treatment, identify possible obstacles to the access and plan the organization of the healthcare network to meet SUS core principles. The qualification of this information is strategic to manage the assistance by SUS.

CONTRIBUTIONS

Adriana Tavares de Moraes Atty and Jeane Glaucia Tomazelli contributed to the study conception and design and data collection. Jeane Glaucia Tomazelli, Adriana Tavares de Moraes Atty, Maria Beatriz Kneipp Dias, Caroline Madalena Ribeiro and Neilane Bertoni contributed for the review, interpretation of the data and wording of the manuscript. Jeane Glaucia Tomazelli, Adriana Tavares de Moraes Atty, Maria Beatriz Kneipp Dias, Caroline Madalena Ribeiro, Arn Migowski and Neilane Bertoni participated of the critical review of the intellectual content of the manuscript and approved the final version.

DECLARATION OF CONFLICT OF INTERESTS

There are no conflict of interests to declare.

FUNDING SOURCES

None.

REFERENCES

1. Ministério da Saúde (BR), Secretaria de Gestão de Investimentos. O cartão nacional de saúde: instrumento para um novo modelo de atenção. Rev Saúde Pública. 2000;34(5):561-4. http://dx.doi.org/10.1590/S0034-89102000000500020
2. Ministério da Saúde (BR); Organização Pan-Americana da Saúde; Fundação Oswaldo Cruz. A experiência brasileira em sistemas de informação em saúde. Vol. 1, Produção e disseminação de informações sobre saúde no Brasil. Brasília, DF: Ministério da Saúde; 2009. (Série B; Textos Básicos de Saúde). [Cap. 3], Sistema de Informações Hospitalares do SUS – SIH-SUS49; p. 49.
3. Ministério da Saúde (BR). Portaria nº 940, de 28 de abril de 2011. Regulamenta o Sistema Cartão Nacional de Saúde (Sistema Cartão). Diário Oficial da União, Brasília, DF; 2011 maio 2. Seção I, p. 58.
4. Ministério da Saúde (BR). Portaria nº 1.560, de 29 de agosto de 2002. Institui o Cartão Nacional de Saúde - Cartão SUS e da outras providências. [Ementa elaborada pela CDI/MS). Diário Oficial da União, Brasília, DF; 2002 set. 13. Seção I, p. 42.
5. Ministério da Saúde (BR). Portaria nº174, de 14 de maio de 2004. Determina que os procedimentos ambulatoriais e hospitalares que exigem autorização prévia sejam acompanhados, obrigatoriamente, da identificação dos usuários por meio do número do Cartão Nacional de Saúde, conforme cronograma constante do anexo desta portaria. [Ementa elaborada pela biblioteca/MS]. Diário Oficial da União, Brasília, DF; 2004 maio 17. Seção I, p. 62.
6. Ministério da Saúde (BR). Portaria nº 2.073, de 31 de agosto de 2011. Regulamenta o uso de padrões de interoperabilidade e informação em saúde para sistemas de informação em saúde no âmbito do Sistema Único de Saúde, nos níveis Municipal, Distrital, Estadual e Federal, e para os sistemas privados e do setor de saúde suplementar. Diário Oficial da União, Brasília, DF; 2011 set. 1. Seção I, p. 63.
7. Ministério da Saúde (BR). DATASUS: relatório executivo da gestão 2011-2014. Brasília, DF: DATASUS; 2015.
8. Ministério da Saúde (BR); Organização Pan-Americana da Saúde; Fundação Oswaldo Cruz. A experiência brasileira em sistemas de informação em saúde. Vol. 1, Produção e disseminação de informações sobre saúde no Brasil. Brasília, DF: Ministério da Saúde; 2009. (Série B; Textos Básicos de Saúde). [Cap. 5], Departamento de Informática do SUS – DATASUS: a experiência de disseminação de informações em saúde; p. 109.
9. Presidência da República (BR). Lei nº 12.527, de 18 de novembro de 2011. Regula o acesso a informações previsto no inciso XXXIII do art. 5º , no inciso II do § 3º do art. 37 e no § 2º do art. 216 da Constituição Federal; altera a Lei nº 8.112, de 11 de dezembro de 1990; revoga a Lei nº 11.111, de 5 de maio de 2005, e dispositivos da Lei nº 8.159, de 8 de janeiro de 1991; e dá outras providências. Diário Oficial da União, Brasília, DF; 2011 nov. 18. Seção I, Edição extra, p. 1.
10. Wickham H, RStudio. Tidyverse: easily install and load the "Tidyverse" [Internet]. Version 1.2.1. Boston: Free Software Foundation, Inc.; 2017 - [cited 2018 Aug 3]. Available from: https://CRAN.R-project.org/package=tidyverse
11. Harron K, Dibben C, Boyd J, et al. Challenges in administrative data linkage for research. Big Data & Society. 2017 Dec;4(2):205395171774567. doi: https://doi.org/10.1177/2053951717745678
12. Coeli CM, Pinheiro RS, Camargo Jr. KR. Conquistas e desafios para o emprego das técnicas de record linkage na pesquisa e avaliação em saúde no Brasil. Epidemiol Serv Saúde. 2015 out.;24(4):795-802. doi: https://doi.org/10.5123/S1679-49742015000400023
13. Jurczyk P, Lu JJ, Xiong L, et al. FRIL: a tool for comparative record linkage. AMIA Annu Symp Proc. 2008 Nov 6:440-4.
14. Zhu Y, Matsuyama Y, Ohashi Y, et al. When to conduct probabilistic linkage vs. deterministic linkage? A simulation study. J Biomed Inform. 2015 Aug;56:80-6. doi: https://doi.org/10.1016/j.jbi.2015.05.012
15. Organisation for Economic Co-operation and Development. Strenghtening heatlh information infrastucture for health care quality governance: good practices, new opportunities and data privacy protection challenges. Paris: OECD Health Policy Studies, OECD Publishing; 2013. doi: http://dx.doi.org/10.1787/9789264193505-en
16. Joint United Nations Programme on HIV/AIDS. Considerations and guidance for countries adopting national health identifiers[Internet]. Geneva, Switzerland: UNAIDS; c2014. [cited 2019 Oct 10]. 112 p. Available from: https://www.unaids.org/sites/default/files/media_asset/JC2640_nationalhealthidentifiers_en.pdf
17. Gadelha MIP, Costa MR, Almeida RT. Estadiamento de Tumores Malignos - análise e sugestões a partir de dados da APAC. Rev Bras Cancerol. 2005;51(3):193-9.
18. Gomes Jr SCS, De Martino R, Almeida RT. Rotinas de integração das tabelas do sistema de autorização de procedimentos de alta complexidade em oncologia do Sistema Único de Saúde. Cad Saúde Coletiva. 2003;11(2):231-54.
19. Souza RC, Freire SM. Integração de dados ambulatoriais de quimioterapia e radioterapia registrados nas bases de dados do SUS. In: 24th Congresso Brasileiro de

Engenharia Biomédica; 2014 out. 13-17. Uberlândia, MG. Rio de Janeiro: Sociedade Brasileira de Engenharia Biomédica; 2014. p. 1904-07.

20. Atty ATM. O câncer de boca no Estado do Rio de Janeiro à luz dos sistemas de informação. [dissertação]. Rio de Janeiro: Escola Nacional de Saúde Pública Sergio Arouca, Fiocruz; 2012.
21. Atty ATM, Tomazelli JG, Dias, MBK. Análise exploratória das informações sobre estadiamento nas autorizações de procedimentos de alta complexidade no Brasil e regiões no período 2010-2014. Rev Bras Cancerol. 2017;63(4):257-64. doi: https://doi.org/10.32635/2176-9745.RBC.2017v63n4.126
22. Teixeira CLS, Bloch KV, Klein CH, et al. Método de relacionamento de bancos de dados do Sistema de Informações sobre Mortalidade (SIM) e das autorizações de internação hospitalar (BDAIH) no Sistema Único de Saúde (SUS), na investigação de óbitos de causa mal-definida no Estado do Rio de Janeiro, Brasil, 1998. Epidemiol Serv Saúde. 2006;15(1):47-57. doi: http://dx.doi.org/10.5123/S1679-49742006000100004.
23. Suzuki KMF. O uso de método de relacionamento de dados (record linkage) para integração de informação em sistemas heterogêneos de saúde: estudo de aplicabilidade entre níveis primário e terciário. [Tese]. Ribeirão Preto, SP: Universidade de São Paulo, Faculdade de Medicina de Ribeirão Preto; 2012.
24. Girianelli VR. Fatores Associados ao Risco de Progressão para Câncer do Colo do Útero ou suas Lesões Precursoras em Mulheres com Exame de Papanicolaou Negativo: Um Estudo de Três Anos de Seguimento. [Tese]. Rio de Janeiro: Instituto Nacional de Câncer; 2008.
25. Cabral MDB. Proposta de relacionamento probabilístico dos registros da base de dados do programa de rastreamento do câncer do colo do útero. [Tese]. Rio de Janeiro: Universidade Federal do Rio de Janeiro; 2010.
26. Tomazelli JG, Girianelli VR, Silva GA. Mulheres rastreadas para câncer de mama: acompanhamento por meio dos Sistemas de Informação em Saúde. Epidemiol Serv Saúde. 2018;27(3): e2017445. doi: http://dx.doi.org/10.5123/s1679-49742018000300005

Recebido em 24/7/2019
Aprovado em 11/10/2019