

Big Data in Population Cancer Research: Challenges and Opportunities

doi: <https://doi.org/10.32635/2176-9745.RBC.2019v65n4.718>

Big Data em Pesquisas Populacionais sobre Câncer: Desafios e Oportunidades

Big Data en la Investigación de Cáncer de Población: Desafíos y Oportunidades

Luiz Claudio Santos Thuler¹; Zulmira Hartz²

INTRODUCTION

In the last years, the popularization of the computers, the increasing access to the Internet, the generation of an astronomic amount of data, actively or passively originated have contributed for the exponential appearance of big databases, the Big Data. They elicit a new challenge which is to identify manners of obtaining and extracting information efficiently and take advantage of the opportunities offered by technological progress¹. Etymologically speaking, the information are originated from data collection, which, on its turn, constitute the base of the knowledge. However, is increasingly difficult to separate data from information since, most of the times, they are interchangeable².

The term Big Data was used for the first time by Michael Cox and David Ellsworth, scientists of the National Aeronautics and Space Administration (NASA), in an article written in 1997, exposing the difficulty of working with very large set of data³. Although there are innumerable definitions of Big Data, what steers us to the reasoning that none of them is satisfactory *per se*, in general, the term has been used when three Vs are present: volume, variety and velocity. These three Vs initially proposed by Laney⁴, have been widely accepted as the main dimensions connected to Big Data. Later, some Vs were added: veracity, versatility and viability and the combination of all Vs would generate value^{1,5,6}. However, the pursue for Vs has not stopped, and Shafer⁷ has recently suggested 42 Vs.

Berman (apud Jackson⁸) proposes that what distinguishes Big Data from Small Data is that the latter are contained in only one computer inside one organization, are generally well structured and formatted (for example, in spreadsheets) and its collection usually

meets a specific goal. The Big Data, not rarely, are distributed through several Internet servers, can be structured or not (for example, free text, images, videos, audios), coming from several sources and generated by many persons, not being designed for a specific purpose of a certain study⁸.

In regard to the generation of health data, Herland et al.⁹ suggest that the use of Big Data should consider two levels: the research question and data collection. The epidemiologic questions, for example, must rely on populational data; the clinical questions must rely on molecular data (as, for instance, the use of data of genic expression to make clinical predictions or subtypes of cancer obtained by anatomopathological exams or yet the use of messages to help patients to obtain medical information), or data collected directly from the patients (for instance, predictive models to identify prognostic factors); on its turn, questions related to human biology can make use of tissue data (lab or imaging exams). In relation to data collection, ideally, it should be considered all the possible sources for its collection. The most used methods in this phase are files, Web trackers and network monitoring software. Next, the data must be processed and stored. In case of data from different sources, they should be primarily integrated and, moving forward, examined for accuracy, completeness and consistency. For its analysis, the metrics, algorithms and forms of visualization must be defined. One of the greatest challenges has been to identify tools that support its storage and management (as, for example: Hadoop Distributed File System – HDFS), volume (Oracle NoSQL, Apache HBase, Apache Cassandra, Apache Hive), processing capacity (MapReduce, Apache Hadoop), capacity of integration (Oracle Big Data Connectors, Oracle Data Integrator) and capacity of statistical analysis (R and Oracle R

¹ National Institute of Cancer José Alencar Gomes da Silva (INCA). Federal University of the State of Rio de Janeiro (Unirio). Rio de Janeiro (RJ), Brazil. Orcid iD: <https://orcid.org/0000-0003-2550-6537>

² Global Health and Tropical Medicine (GHTM). Institute of Hygiene and Tropical Medicine (IHMT) of the New University of Lisbon (UNL). Lisbon, Portugal. Orcid iD: <https://orcid.org/0000-0001-9780-9428>

Address for correspondence: Luiz Claudio Santos Thuler. Division of Clinical Research of INCA. Rua André Cavalcanti, 37 – Rio de Janeiro (RJ), Brazil. CEP 20231-050. E-mail: lthuler@gmail.com



Enterprise). Finally, based in the results, decisions can be taken to reinforce positive behaviors or change undesired behaviors¹.

BIG DATA AND CANCER

Several authors have been utilizing the Big Data to respond to specific questions about cancer. Meyer et al.¹⁰ propose the use of Big Data for the research of populational base in cancer and quote several authors that have used The Integrated Cancer Information and Surveillance System – ICISS of North Carolina, USA and the health insurance system Medicare. These authors emphasize, for example, the article of Wheeler et al.¹¹ that discusses the inequalities of the utilization of radiotherapy after conservative surgery in low-income women with breast cancer in initial stage. In this study, only 75% of the women initiated radiotherapy within six months after the diagnosis of breast cancer and, although the offer of radiotherapy have improved during the years, its beginning within one year was more frequent in patients with less comorbidities and in chemotherapy. The study of Holmes et al.¹² also presented by Meyer et al.¹⁰, showed that patients who lived farther away from the urologist were more prone to be diagnosed with high risk prostate cancer. The authors emphasize that, disproportionally, the distance impacted more negatively African-descendants than Caucasian men: among those who lived more than 32 km away from the closest urologist, 49% of the Caucasian patients and 79% of the African-descendant patients were diagnosed with disease of high risk.

In an article about the use of Big Data to evaluate the quality of care in oncology, Broughman et. al.¹³ propose that, to be considered of high quality, the services must be safe, effective, patient focused, timely, efficient and equitable. The authors cite 26 studies that utilize the data of The Surveillance, Epidemiology, and End Results - SEER of the USA National Institute of Cancer and the National Cancer Database – NCDB with that finality. In these studies, the number of patients varied between 1,962 and 2,342,558, being highlighted, for instance, the study of Fedewa et al.¹⁴, that analyzed 312,339 men for prostate cancer severity at the diagnosis. African American, Hispanic and Asian patients, in comparison with Caucasian and those uninsured and beneficiary of federal and state social programs with medical costs covered for individuals with limited resources and income (Medicaid), in comparison with private insured patients, were more prone to present more aggressive disease at the diagnosis. Similarly, they quote the study elaborated by Gross et al.¹⁵, with data from SEER and Medicare to show that African American patients versus Caucasian had less likelihood of receiving treatment for lung cancer (surgery)

breast (radiotherapy), colon (adjuvant therapy) and prostate (prostatectomy, brachytherapy or radiotherapy).

The major European public-private initiative, The Innovative Medicines Initiative 2 - IMI2, a joint undertaking between the European Union and the European Federation of Pharmaceutical Industries and Associations – EFPIA, launched recently the program Big Data for Better Outcomes - BD4BO. The program aims to catalyze and support the evolution of outcome-driven sustainable healthcare systems. In cancer, the main goal is to increase the body of evidences to improve the outcomes. Through epidemiologic, clinical, economic measures (Patient Reported Outcomes – PRO) it will be analyzed the predictive factors of complications and risk of development of adverse events through all the stages of the disease¹⁶.

BIG DATA AND CANCER IN BRAZIL

In Brazil, the information about cancer have been obtained from different sources according to its specificity.

The System of Information about Mortality (SIM) has available data about deaths by all causes, including cancer since 1979 and has been widely utilized, particularly to analyze the trends of mortality by different types of cancer in the country. Based in these data, indicators and rates of mortality by cancer in Brazil are provided by the National Institute of Cancer José Alencar Gomes da Silva (INCA)¹⁷ as tables, graphs, and maps in the Mortality Online Atlas.

Data about the incidence of cancer can be obtained from the Population-Based Cancer Registries (PBCR) that has the objective of collecting data about all the cancer incident cases residing in a limited geographic area occurred in a certain period. The first PBCR were implemented in Recife (1967) and in São Paulo (1969)¹⁸. Currently, there are more than 20 PBCR in activity with at least five years of consolidated information, covering nearly ¼ of the population of the country¹⁹. Its data have been utilized in the analysis of the incidence of several types of cancer and in specific populations.

On its turn, the Hospital-based Cancer Registry (HBCR) concentrate data of the patients enrolled in only one hospital, being useful to evaluate the quality of the care provided. In Brazil, the first HBCR was created in 1982 in INCA, initiating data collection from January 1983¹⁹. Currently, a data tabulator (*Integrador RHC*) offers at the Internet an assemblage of more than three million cases of cancer, covering the 27 Federation Units. The data concentrate between 2000 and 2017 (99.9% of the cases) and come from more than 330 hospitals, with number of cases per hospital varying from 17 to 158,516 (data updated on October 29, 2019). Several studies have utilized the data of *Integrador RHC* to respond to

questions involving since the risk factors for cancer until its treatment and outcomes.

Information about cytopathological, anatomopathological and mammographies performed by the Public Health Care System (SUS) are available by the Cervical Cancer Information (SISCOLO) and by the Breast Cancer Control Information System (SISMAMA), respectively. The data of SISCOLO and SISMAMA, joined recently in SISCAM have been used to monitor the guidelines, objectives, goals and municipal, state, regional and national indicators of SUS agreed upon in the tripartite inter-management commission²⁰. The two specific indicators for cancer are the ratio between the number of cytopathological exams of the cervix in women from 25 to 64 years old and the population of the same age range, and the ratio between the number of exams of screening mammography made in women from 50 to 69 years old and the population of the same age range. The edition of 2015 (last available in October 2019) presents a historical series from 2008 to 2015, allowing to follow the trend of each indicator in the territory analyzed.

There are still the unspecified databases for cancer, kept by the Ministry of Health, as, for example, the Live Birth Information System (SISNAC), the Notifiable Diseases Information System (SINAN), the Primary Healthcare Information System (SIAB), the Outpatient Information System (SIA), the Hospital Information System (SIH), the High Complexity Procedures Authorization (APAC) and the databases of investigation and researches available at the website of the Information Technology Department of the Public Health Care System (DATASUS) and that can be useful according to the object of the study²¹.

Although these databases do not necessarily meet all the desirable dimensions of the Big Data as velocity (some data are provided between one and three years after its occurrence), other characteristics as volume, veracity, versatility, viability, overall, are present.

EXAMPLES OF THE USE OF BIG DATA IN CANCER IN BRAZIL

In a pioneer study with “*Integrador RHC*”, Thuler et al.²², after analyzing 77,317 cases of cervical cancer, indicated the elevated percent of cases whose diagnosis occurred when the disease was already in advanced stage. More than 1/3 of the cases presented in stages III and IV at the moment of the diagnosis. The authors emphasize yet that the percent of the cases with advanced stage increased from 32.8% to 35.5% between the first (2000-2004) and the second (2005-2009) quinquennium analyzed. In another study, also with data of “*Integrador RHC*”, this time involving 59,317 women with breast cancer, Abrahão et al.²³ identified factors associated to the diagnosis in stage ≥IIB. The adjusted analysis showed that <50 years old,

low education and living in the country’s poorest regions were associated to advanced disease at the moment of the diagnosis.

The incidence and the mortality of lung cancer between 2000 and 2014 were analyzed from three databases: PBCR, “*Integrador RHC*” and SIM, respectively. Costa et al.²⁴ showed that, while the incidence rates of mortality by lung cancer dropped for males, they increased for females. They draw attention for the reduction of the ratio man:woman during the period. Using the same databases, Melo et al.²⁵ performed the biggest study ever made in Brazil about cutaneous melanoma. They verified that the increase of the incidence rates observed for both genders in the period was followed by rates of mortality practically stable, indicating possible improvements in the diagnosis and treatment of the cases. Still, they highlighted the fact that nearly half of the patients had the diagnosis in advanced stages of the disease.

The big national databases have also been used to respond to frequent questions about child and youth cancer. In a study with the objective of analyzing the trends of cervical cancer in adolescents and young women (15 to 29 years), Viana et al.²⁶ utilized data of 21 PBCR and concluded that, in Brazil, the incidence rates are elevated and are growing or stable for the most part of the PBCR, showing the necessity of mandatory interventions targeted to the prevention and control of the disease in this specific age range. Additionally, Reis et al.²⁷ utilized data of 12 PBCR to analyze the risk of developing leukemias in early childhood (children ≤5 years old). When proposing a cohort-case study, 272 patients with leukemia were compared to 1,088 controls identified from the database of SISNAC in the same age range, location and period (2000-2009). It was observed in children with congenital mal formation and in those whose mothers worked exposed to chemical products (agriculture, chemical and petrochemical industry), higher risk of developing leukemia.

The standard of cervical abnormalities in the State of Maranhão was analyzed by Pinho-França et al.²⁸ with data from SISCOLO. The evaluation of the results of cervical cytology of 1,363,689 women examined between 2007 and 2012 revealed that as worst the Human Development Index (HDI) of the municipality where the woman lives, higher is the risk of presenting high-grade squamous intraepithelial injury or cervical cancer. The authors emphasized that, to control the disease, it is necessary to reduce social inequalities and improve the offer of health services. In the other hand, analyzing the association of HDI and the stage of cervical cancer at the diagnosis, Vale et al.²⁹ showed that, from 51,158 cases registered in “*Integrador RHC*”, there is a reverse correlation between

HDI of the city where the woman lives and the stage of the disease. The authors draw attention for the elevated percent of cases diagnosed in advanced stages and indicated the necessity of regional strategies that improve the efficiency of the interventions targeted to cancer in countries with great socioeconomic disparities as Brazil.

When Law number 13,685³⁰ came into force on December 24, 2018, the notification of neoplasms-associated diseases, injuries and events attended in public and private health facilities started to be mandatory; this disposition widened the perspectives of having complete cancer data in the country. However, this law still needs to be regulated.

CHALLENGES AND OPPORTUNITIES OF USING BIG DATA IN POPULATIONAL STUDIES ABOUT CANCER

Although it is envisaged a huge potential of utilization of Big Data in the health area, there are peculiarities to be considered. The interpretation of the results of statistical tests with great volumes of data must be cautious as the likelihood of identifying associations as statistically significant increases, but that, often, cannot be defined as causal or clinically important associations^{31,32}. On the other hand, the quantity of existing variables in the databases widens the odds of identifying spurious associations obtained from disassociated analyzes from logical and/or theoretical models. It is being recommended as strategies to avoid to perform tests of hypothesis for all the variables available, limiting only to those for which there is theoretical substantiation, using strategies to reduce the occurrence of error type I (when the result of the test has statistical significance, but, actually its occurrence happened haphazardly) as the test of Bonferroni, the correction of Bonferroni and the control of the false discoveries rates through the procedures of Benjamini-Hochberg^{21,33}. Other issues that cannot be left apart address the validity of the construct and reliability³⁴. In this direction, further to the necessary dimensions of the Big Data more frequently described as velocity, volume, variety, veracity, versatility, viability, it is necessary the V of validity proposed by Normandeau³⁵. It is worth mentioning that, once utilized with finality of research, the Big Data are not exempt from complying with the ethical demands of researches involving human subjects, beings, being necessary to ensure privacy, confidentiality, safety and integrity of the data^{21,36}.

CONCLUSION

In this context, considered the possibilities of using the Big Data in populational researches about cancer, it is essential to grant utility and visibility to data about cancer offered by public agencies.

CONTRIBUTIONS

Both authors contributed equally and substantially in all phases of the work and approved the final version for publication.

DECLARATION OF CONFLICT OF INTERESTS

There is no conflict of interests to declare.

FUNDING SOURCES

None.

REFERENCES

1. Singh MK, Kumar DG, editors. Effective big data management and opportunities for implementation [Internet]. Pennsylvania, USA: IGI Global; 2016. Chapter 1, Big data: challenges, opportunities and realities; [cited 2019 Sept 17]; p. 1-24. Available from: <https://arxiv.org/ftp/arxiv/papers/1705/1705.04928.pdf>
2. The Economist [Internet]. London: The Economist Newspaper; [1997]. Data, data everywhere; [2010 Feb 27]. [cited 2019 Sept 17]. Available from: <https://www.economist.com/special-report/2010/02/27/data-data-everywhere>.
3. Cox M, Ellsworth D. Managing big data for scientific visualization. ACM Siggraph [Internet]. 1997 [cited 2019 Sept 17];97:21-38. Available from: https://www.researchgate.net/publication/238704525_Managing_big_data_for_scientific_visualization
4. Laney D. 3D data management: controlling data volume, velocity, and variety [Internet]. Stamford: META Group; 2001 Feb 6. [cited 2019 Sept 17]. File: 949. Available from: <https://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf>
5. Dias JA, Duarte P. Big data opportunities in healthcare. How can medical affairs contribute? *Rev Port Farmacoter*. 2015;7(4):230-236. doi: <https://doi.org/10.25756/rpf.v7i4.99>
6. Magalhães J, Hartz Z, Martins MRO. Big Data para a investigação em saúde e a ciência aberta: um contributo para a gestão do conhecimento. *An Inst Hig Med Trop*. 2016;15(Supl. 2):S75-S82. doi: <https://doi.org/10.25761/anaisihmt.130>
7. KDnuggets [Internet]. [place unknown]: KDnuggets; 1997. Shafer T. The 42 V's of big data and data science; 2017. [cited 2019 Sept 17]. Available from: <https://www.kdnuggets.com/2017/04/42-vs-big-data-data-science.html>
8. Jackson S. Big data monitoring and evaluation: a theoretical framework, tools, and lessons learned from practice

- [Internet]. [Draft v2.0] Jakarta: United Nations Global Pulse; 2015 Dec. [cited 2019 Sept 17]. Available from: http://unglobalpulse.org/sites/default/files/Annex%201%20Big_data_monitoring_and_evaluation.pdf
9. Herland M, Khoshgoftaar TM, Wald R. A review of data mining using big data in health informatics. *J Big Data*. 2014;1(2). doi: <https://doi.org/10.1186/2196-1115-1-2>
 10. Meyer AM, Olshan AF, Green L, et al. Big data for population-based cancer research: the integrated cancer information and surveillance system. *N C Med J*. 2014 Jul-Aug;75(4):265-9. doi: <https://doi.org/10.18043/ncm.75.4.265>
 11. Wheeler SB, Wu Y, Meyer AM, et al. Use and timeliness of radiation therapy after breast-conserving surgery in low-income women with early-stage breast cancer. *Cancer Invest*. 2012;30(4):258-267. doi: <https://doi.org/10.3109/07357907.2012.658937>
 12. Holmes JA, Carpenter WR, Wu Y, et al. Impact of distance to a urologist on early diagnosis of prostate cancer among black and white patients. *J Urol*. 2012;187(3):883-888. doi: <https://doi.org/10.1016/j.juro.2011.10.156>
 13. Broughman JR, Chen RC. Using big data for quality assessment in oncology. *J Comp Eff Res*. 2016;5(3):309-19. doi: <https://doi.org/10.2217/cer-2015-0021>
 14. Fedewa SA, Etzioni R, Flanders WD, et al. Association of insurance and race/ethnicity with disease severity among men diagnosed with prostate cancer, National Cancer Database 2004-2006. *Cancer Epidemiol Biomarkers Prev*. 2010 Oct;19(10):2437-44. doi: <https://doi.org/10.1158/1055-9965.EPI-10-0299>
 15. Gross CP, Smith BD, Wolf E, et al. Racial disparities in cancer therapy: did the gap narrow between 1992 and 2002?. *Cancer*. 2008 Feb 15;112(4):900-8. doi: <https://doi.org/10.1002/cncr.23228>
 16. Szócska M, Jayawardana S, Smand C, et al. Big data for better outcomes: supporting health care system transformation in Europe. *Eurohealth* [Internet]. 2017. [cited 2019 Sept 17];23(1):7-9. Available from: http://www.euro.who.int/__data/assets/pdf_file/0004/335812/Eurohealth-V23-N1-2017.pdf?ua=1
 17. Instituto Nacional de Câncer José Alencar Gomes da Silva. Atlas on-line de mortalidade [Internet]. Rio de Janeiro: INCA; [2014] - [atualizado 2019 maio 30; acesso 2019 set. 17]. Disponível em: <https://mortalidade.inca.gov.br/MortalidadeWeb/>
 18. Ministério da Saúde (BR), Departamento de Ciência e Tecnologia, Secretaria de Ciência e Tecnologia e Insumos Estratégicos. Integração de informações dos registros de câncer brasileiros. *Rev Saúde Publica*. 2007 Oct;41(5):865-8. doi: <http://dx.doi.org/10.1590/S0034-89102007000500024>
 19. Instituto Nacional de Câncer José Alencar Gomes da Silva. Informação dos registros hospitalares de câncer como estratégia de transformação: perfil do Instituto Nacional de Câncer José Alencar Gomes da Silva em 25 anos [Internet]. Rio de Janeiro: INCA; 2012. [acesso 2019 set. 17]. 100 p. Disponível em: <https://www.inca.gov.br/publicacoes/livros/informacao-dos-registros-hospitalares-de-cancer-como-estrategia-de-transformacao>
 20. Ministério da Saúde (BR); Comissão Intergestores Tripartite. Resolução nº 5, de 19 de junho de 2013. *Diário Oficial da União, Brasília, DF*; 2013 jun. 20. Seção I, p. 63.
 21. Chiavegatto Filho ADP. Uso de big data em saúde no Brasil: perspectivas para um futuro próximo. *Epidemiol Serv Saúde*. 2015;24(2):325-332. doi: <http://dx.doi.org/10.5123/S1679-49742015000200015>
 22. Thuler LCS, Bergmann A, Casado L. Perfil das pacientes com câncer do colo do útero no Brasil, 2000-2009: estudo de base secundária. *Rev Bras Cancerol*. 2012;58:351-357. doi: <https://doi.org/10.32635/2176-9745.RBC.2012v58n3.583>
 23. Abrahão KS, Bergmann A, Aguiar SS, et al. Determinants of advanced stage presentation of breast cancer in 87,969 Brazilian women. *Maturitas*. 2015 Dec;82(4):365-70. doi: <https://doi.org/10.1016/j.maturitas.2015.07.021>
 24. Costa GJ, Mello MJG, Ferreira CG, et al. Increased incidence, morbidity and mortality rates for lung cancer in women in Brazil between 2000 and 2014: an analysis of three types of sources of secondary data. *Lung Cancer*. 2018 Nov;125:77-85. doi: <https://doi.org/10.1016/j.lungcan.2018.09.005>
 25. Melo AC, Wainstein AJA, Buzaid AC, et al. Melanoma signature in Brazil: epidemiology, incidence, mortality, and trend lessons from a continental mixed population country in the past 15 years. *Melanoma Res*. 2018;28(6):629-36. doi: <https://doi.org/10.1097/CMR.0000000000000511>
 26. Viana LS, Balmant NV, Silva NP, et al. Incidence Trends of Cervical Cancer in Adolescents and Young Adults: Brazilian Population Based Data. *J Adolesc Young Adult Oncol*. 2018 Feb;7(1):54-60. doi: <https://doi.org/10.1089/jayao.2017.0048>
 27. Reis RS, Silva NP, Santos MO, et al. Mother and child characteristics at birth and early age leukemia: a case-cohort population-based study. *J Pediatr*. 2017 Nov-Dec;93(6):610-18. doi: <http://dx.doi.org/10.1016/j.jpeds.2016.12.009>
 28. Pinho-França JR, Chein MBC, Thuler LCS. Patterns of cervical cytological abnormalities according to the Human Development Index in the northeast region of Brazil. *BMC Womens Health*. 2016 Aug 12;16:54. doi: <http://dx.doi.org/10.1186/s12905-016-0334-2>
 29. Vale DB, Sauvaget C, Muwonge R, et al. Level of human development is associated with cervical cancer stage at diagnosis. *J Obstet Gynaecol*. 2019;39(1):86-90. doi: <http://dx.doi.org/10.1080/01443615.2018.1463976>

30. Presidência da República (BR). Lei nº 13.685, de 25 de junho de 2018. Diário Oficial da União, Brasília, DF; 2018 jun. 26. Seção I, p. 1.
31. Groeneveld PW, Rumsfeld JS. Can big data fulfill its promise? *Circ Cardiovasc Qual Outcomes*. 2016;9(6):679-82. doi: <http://dx.doi.org/10.1161/CIRCOUTCOMES.116.003097>
32. Kamper SJ. Interpreting Outcomes 2-Statistical Significance and Clinical Meaningfulness: Linking Evidence to Practice. *J Orthop Sports Phys Ther*. 2019 Jul;49(7):559-60. doi: <http://dx.doi.org/10.2519/jospt.2019.0704>
33. McDonald JH. Handbook of Biological Statistics [Internet]. 3rd ed. Baltimore: Sparky House Publishing; 2014 [revised 2015 July 20]. Maryland, multiple comparisons; [cited 2019 Sept 17]; [about I screen]. Available from: <http://www.biostathandbook.com/multiplecomparisons.html>
34. Lazer D, Kennedy R, King G, et al. The parable of google flu: traps in big data analysis. *Science* [Internet]. 2014 [cited 2019 Sept 17];343:1203-1205. Available from: <https://gking.harvard.edu/files/gking/files/0314policyforumff.pdf>
35. insideBIGDATA [Internet]. Portland, OR: insideBIGDATA; [2011]. Normandeau K. Beyond volume, variety and velocity is the issue of big data veracity; 2013 Sept 12. [cited 2019 Sept 17]. Available from: <https://insidebigdata.com/2013/09/12/beyond-volume-variety-velocity-issue-big-data-veracity/>
36. Salerno J, Knoppers BM, Lee LM, et al. Ethics, big data and computing in epidemiology and public health. *Ann Epidemiol*. 2017;27(5):297-301. doi: <http://dx.doi.org/10.1016/j.annepidem.2017.05.002>

Recebido em 30/10/2019
Aprovado em 1/11/2019