Design of a geospatial model applied to Health management

Concepção de um modelo geoespacial aplicado na gestão à saúde

La producción de un modelo geoespacial aplicado a la gestión de la salud

Marcelo Rosano Dallagassa[I,II]
ORCID: 0000-0003-0570-4405

Franciele Iachecen[II]
ORCID: 0000-0002-2019-1479

Deborah Ribeiro Carvalho[I]
ORCID: 0000-0002-9735-650X

Sergio Ossamu Ioshii[I]
ORCID: 0000-0003-1208-2448

[I]*Pontifícia Universidade Católica do Paraná. Curitiba-PR, Brasil.*
[II]*Unimed Federação do Estado do Paraná,*
Core of Health Intelligence. Curitiba-PR, Brasil.

Corresponding Author:
Marcelo Rosano Dallagassa
E-mail: dallagassa@unimedpr.coop.br

ABSTRACT
Objective: To identify geographically the beneficiaries categorized as prone to Type 2 Diabetes Mellitus, using the recognition of patterns in a database of a health plan operator, through data mining. **Method:** The following steps were developed: the initial step, the information survey. Development, construction of the process of extraction, transformation, and loading of the database. Deployment, presentation of the geographical information through a georeferencing tool. **Results:** As a result, the mapping of Paraná according to its health care network and the concentration of Type 2 Diabetes Mellitus is presented, enabling the identification of cause-and-effect relationships. **Conclusion:** It is concluded that the analysis of georeferenced information, linked to health information obtained through the data mining technique, can be an excellent tool for the health management of a health plan operator, contributing to the decision-making process in Health.
Descriptors: Health Care; Data Mining; Geographic Mapping; Supplementary Health; Chronic Disease.

RESUMO
Objetivo: Identificar geograficamente os beneficiários categorizados como propensos à doença Diabetes Mellitus Tipo 2, utilizando o reconhecimento de padrões em uma base de dados de uma operadora de plano de saúde, por meio da mineração de dados. **Método:** Desenvolveram-se as seguintes etapas: fase inicial, levantamento de informações. Desenvolvimento, construção do processo de extração, transformação e carga do banco de dados. Implantação, apresentação das informações geográficas por meio da ferramenta de georreferenciamento. **Resultados:** Como resultados, apresenta-se o mapeamento do Paraná em relação a sua rede assistencial e a concentração de Diabetes Mellitus Tipo 2, oportunizando a identificação de relações de causa-efeito. **Conclusão:** Conclui-se que a análise de informações georreferenciadas, vinculadas às informações de saúde obtidas por meio da técnica de mineração de dados, pode ser um excelente instrumento para a gestão da saúde de uma operadora de plano de saúde, contribuindo para o apoio à tomada de decisões em saúde.
Descritores: Atenção à Saúde; Mineração de Dados; Georreferenciamento; Saúde Suplementar; Doenças Crônicas.

RESUMEN
Objetivo: Identificar geográficamente a los beneficiarios categorizados como propensos a la enfermedad Diabetes mellitus tipo 2, utilizándose el reconocimiento de patrones en una base de datos de cierta compañía de seguro médico por medio de la minería de datos. **Método:** Se desarrollaron las siguientes etapas: fase inicial, levantamiento de información. Desarrollo, construcción del proceso de extracción, transformación y carga en la base de datos. Implantación, presentación de la información geográfica mediante la herramienta de georreferenciación. **Resultados:** Se presenta el mapeo de Paraná (Brasil) con relación a su red asistencial y la concentración de Diabetes mellitus tipo 2, proporcionando la identificación de las relaciones de causa-efecto. **Conclusión:** Se concluyó que el análisis de las informaciones georreferenciadas, vinculadas a las informaciones de salud obtenidas por la técnica de minería de datos, puede ser una excelente herramienta en la gestión de salud de cierta compañía de seguro médico, lo que contribuye al apoyo a la toma de decisiones en salud.
Descriptores: Atención a la Salud; Minería de Datos; Mapeo Geográfico; Salud Suplementaria; Enfermedades crónicas.

INTRODUCTION

Epidemiology is the study of interrelationships between several determinants of disease frequency and distribution within a population. Such knowledge is fundamental in preventing and treating diseases, offering consistent elements for referrals in the Health field. The identification of patterns on the occurrence of diseases in human populations, in addition to the factors that influence and condition them, define the study object of Epidemiology[1].

In London, 1854, the number of cholera cases was regarded as a stable and low incidence; nonetheless, at a given time, cholera became a major epidemic, coming to register more than 500 fatal cases within about 12.5 hectares in a 10-day period. John Snow, considered by many as the father of Epidemiology, developed at the time an observational study linking the cases geographically, identifying the midpoint – a water spout in the neighborhood of *Soho* – as being responsible for the spread of the disease. Through the cause-and-effect relationship, after careful and intensive investigations, he concluded the other hypotheses about the origin of the disease should be rejected, claiming that the water route was responsible for the transmission of *Vibrio cholerae*[1].

Another example of the association of geographical positioning with a cause-and-effect relationship is the Li-Fraumeni syndrome, a pathology that makes individuals more vulnerable to certain types of neoplasms, identified in the mutation of gene expression. It was observed these cases were usually located in the South and Southeast regions of Brazil and, particularly, the geographical points of the individuals with the syndrome were bound to the path of a Portuguese immigrant in the 18th century – today, a proven theory of the genetic offspring of the disease[2].

Currently, Brazil is going through a demographic transition due to the decreased infant mortality and increased life expectancy of the population, which went from 50 to 73 years, on average. Other factors also influence this scenario, e.g., the decreased fertility rate among women, which in the 1960s was on average six children and currently is less than two. These factors have a direct impact on the future national scenario, and it is estimated there will be 64 million older people in 2050, representing 29.7% of the Brazilian population[3].

With the increased life expectancy, one can see a change in the epidemiological scenario – the growth of non-communicable chronic diseases in the general population[4]. Population aging has a direct impact on health spending, driven by the increasing proportion of older adults in the population and the growing use of assistance resources. The magnitude of the increase in health costs will depend on the people's quality of life and on the existence or not of diseases and comorbidities[3]. Before these scenarios, the World Bank underlines the importance of organizing the health systems to adapt to the new epidemiological profile, stating that health promotion and disease prevents will continue to be the greatest challenges for the sector.

In the field of supplementary health insurance, there is a challenge for health plans regarding their sustainability. This task has been hampered due to legal constraints posed by the National Agency of Supplementary Health (ANS), which impact directly on their costs due to the fact the limitation of contract adjustments exceed the medical-hospital inflation Variance of

Medical-Hospital Costs (VCHM) that, in December 2016, in the period of 12 months, had a buildup of 20.4%[5], an index much higher than the general inflation index Extended Consumer Price Index (IPCA), which for the same period was 6.29%[6]. In addition, every two years there is the introduction of new and expensive technologies and procedures that extend the list of mandatory coverage services for the operators of health plans.

Given this context, it is necessary to use applications that enable quick and intelligent measures to minimize the direct costs of the operation, seeking a balance between customer satisfaction and the service providers since "guaranteeing the minimum solvency conditions of insurers is essential to enable the existence of an insurance market that meets the objective of protecting the interests of insured persons"[7].

However, health plan operators encounter difficulties to act in the prevention of chronic diseases. One of the reasons for this is directly associated with the lack of beneficiaries' clinical data systematized in their databases, making it impossible to extract information and knowledge in a more automated way. This is also due to the fact the construction of information systems within these companies is directed only at the administrative control, aiming only at the payment of service providers and the management of contracts[8].

In 2007, Resolution No. 1,819 of the Federal Council of Medicine[9] was published, prohibiting the placing of the International Classification of Diseases (ICD) on completing the Supplementary Health Information Exchange guidelines (TISS) in ambulatory care.

The great amount of information generated by health information systems, added to information from external environments that are often fed in real time and made available in different formats (texts, videos, images, messages, gene expressions etc.), define the term "Big Data"[10].

The discovery of patterns in the "Big Data" environment using traditional methods of analysis, besides demanding a lot of time and resources, does not guarantee the full exploitation of its potentials[10]. With the use of data mining, a non-trivial process for finding hidden and possibly useful information, this work becomes more efficient, enabling the support to decision-making processes[11].

Integrating health information, linked to geographic, environmental, socioeconomic, and demographic data, allows the creation of hypotheses for scientific research on the causes and origins of certain diseases, providing knowledge on the prevalence, incidence, transmission control, and treatment of diseases. The analysis of georeferenced information is also of great importance in the discovery of cause-and-effect relationships, providing a dynamic analysis and enabling the identification of more vulnerable groups and the knowledge on the current health status of the population[12-13].

There are numerous databases that, when aggregated and enriched with other information linked to a geographical location record, allow the discovery of important characteristics for the identification of patterns in a given region.

Linking pattern recognition methodologies with georeferencing tools to a well-formulated research question can generate useful tools for the discovery of knowledge in the health area, enabling the implementation of important practices to reduce

costs and improve the population quality of life, even considering the difficulties related to the absence of clinical information.

Thus, the question this research intends to address arises from the difficulties and needs identified. What are the techniques and methods of knowledge discovery that enable the creation of geographically referenced indicators for monitoring the health care of the group of beneficiaries of a health plan operator, aimed at health promotion and disease prevention?

Some related works were identified for the recognition of diseases through the database, which consider the use of associated procedures, algorithms, and methods that enable identifying in the accounts payment records the discovery of patterns of use linked to certain diseases, among which one can cite a study in the region of Pávia, Italy, which consisted of research with administrative and clinical data on diabetes, using rules of temporal association. The research aimed at analyzing data from the health system of the region. The method also highlighted the frequent temporal associations of interest in the diagnosis related to the patient's clinical condition[14].

Another study, also developed at the University of Pavia, Italy, consisted of a method to identify patterns based on temporal data, which were extracted as rules of the temporal association. This research concluded that there is much potential for data mining in searching temporal association rules, suggesting that these methods can be more exploited since the demand for tools that uncover potential rules of interest for managers is increasing[15].

One can also mention another methodology for identifying beneficiaries in a health plan operator with Type 2 Diabetes Mellitus indicatives in the state of Paraná, Brazil. By a history of use, the selection of relevant variables for data generation was carried out. The selection was submitted to the algorithm J48 for finding rules, later evaluated by a group of experts. From this technique, one could extract the patterns for other chronic diseases, which are now part of applications for identifying and categorizing beneficiaries[8].

From this problematization and the definition of these concepts, the need arises for constructing an environment that aggregates information from diverse sources for the enrichment of the internal databases, integrated with pattern recognition techniques and associated to a geographical reference, allowing the identification of diseases that affect the population to promote a more effective monitoring of chronic non-communicable diseases and to enable the development of regional actions of health promotion programs.

OBJECTIVE

To identify geographically the beneficiaries categorized as prone to Type 2 Diabetes Mellitus (DM), using the recognition of patterns in a database of a health plan operator, through data mining.

METHOD

Ethical aspects

This research was carried out with the acquiescence of the institution and project submission to the Research Ethics Committee of the Pontifical Catholic University of Paraná (PUCPR).

Study design, location, and period

This was a descriptive study of a quantitative approach, retrospective in nature, using the database of a large operator of health plans in the State of Paraná, in 2017.

Population or sample: inclusion and exclusion criteria

The sample was based on beneficiaries of the health insurance plan of the operator, who were active at the period observed and who, due to their use and frequency of services used, were categorized as prone to Type 2 DM. As inclusion criterion, the selection of active health plan beneficiaries in 2017, who fell into the category of prone to Type 2 DM, was established; the exclusion criteria were inactive beneficiaries in 2017, and who were not categorized for Type 2 DM. The choice for this pathology is due to this being a constantly-growing disease, stemmed from several factors, such as obesity, sedentary lifestyle, population aging, and increased survival rate of patients with the disease[8]. In addition, it is estimated that non-communicable chronic diseases prompted 38 million deaths in 2012; only Type 2 DM was the root cause for 5.3% of them[16].

Study protocol

After structuring the database, the study was divided into three stages: the initial stage, development, and deployment. In the initial stage, a survey of the possible bases of public research on health records was carried out with the help of the institution's sectors, aimed a constructing the theoretical framework and identifying and justifying possible applications and opportunities for the development of the project, with due approval by the managers of the organization. Furthermore, in planning and in conjunction with managers, the possible databases with areas of the organization were listed. Listed bases were obtained through a questionnaire. The research was formulated by the authors of the research, with the purpose of raising the required databases to compose the project. Such questionnaire was applied on October 3rd, 2016, at the Head office of the health plan operator (HPO). Prior to the questionnaire application, a presentation was held, discussing the main objective of the project, as well as the public bases already covered. Then, 27 questionnaires were delivered and, of them, ten (41.6%) were returned completed. The research had as target audience experts who work in the sectors of information of the HPO in different cities in the State of Paraná. The questionnaire was structured in three questions: one referring to the field of Health, another to the Market, and the latter to the strategic area. Each question contained an assessment on the importance of each area, which ranged from 0 to 5 – with 0 standing for little importance, 3, average importance, and 5 referring to very important; moreover, an open question was proposed, requiring interviewees to describe other possible information sources to be applied. Still in this stage, a tool was defined to be used in elaborating the process of development and deployment, with the premise of using a free software. In the development stage, the software previously listed was installed and a training on the software applications Quantum GIS[17] and Pentaho PDI[18] was applied to the internal team responsible for

the project; then, the step of environment preparation started, disposing the data of beneficiaries categorized as prone to Type 2 DM according to city of dwelling. Finally, the integration of information of interest obtained with the questionnaire results in georeferenced layers was started, with the construction of a panel for viewing this information geographically. In the deployment stage, the activity was the ratification of the tool, with the presentation of results obtained and the mapping of information layers, in which the more intense the color the greater the number of beneficiaries prone to Type 2 DM.

Analysis of results and statistics

Data were organized into tables, for developing studies and discussing them according to the literature available on the topic. For developing georeferenced layers, the georeferencing software Q-GIS was used. Data contained in the table were transferred to this software through a junction with a statewide layer with the 399 municipalities in Paraná, obtained on the IBGE website. The results of the beneficiaries, categorized through pattern recognition using data mining, with Type 2 DM were disposed according to the city within the Paraná state. The prevalence of each municipality was calculated based on the population who had a health plan – the numerator was the Type 2 DM and, as the denominator, the number of beneficiaries in this period, multiplied by 1,000.

RESULTS

Initial stage – Theoretical framework and planning

In the planning activity, through interviews and meetings with the sectors of the organization, some opportunities were identified that could be useful for composing the geographical database, including the following sources: ANS[19], MS (Ministry of Health)[20], SESAPR (State Secretariat of Health of Paraná)[21], PMC (City Hall of Curitiba) [22], Datasus (Department of Informatics of SUS) – Tabnet (Health Information)[23], DW, Simepar (Meteorological System of Paraná)[24], Satisfaction and Use Surveys, among others. The idea of choosing these databases consists of composing the information relating to epidemiological profile, health care network, health technology assessment, data on climate, health guidelines and protocols on the quality of care.

In the questionnaire application, the following results were achieved: in the field of Health, 100% of the interviewees scored 5 (very important). In the area of Market, 80% of the interviewees scored 5 (very important), 10% considered of average importance and 10% did not respond. In the area of Strategy, 80% of the interviewees scored 5 (very important) and 20% did not respond.

In the open questions, the following results were obtained: in the area of Health, were cited as relevant sources of application to the study: knowledge on the dwelling location, sanitation, access to goods and services, cellular network, vaccine coverage indicators, research of temporary partial coverage with beneficiaries, health applications, mortality, morbidity and birth rates. In the area of Market, were cited: demographic density of municipalities, mapping of industries and companies, the population of the cities, the population of beneficiaries, and commercial associations. In the

strategic area, they cited: knowledge on the greater movement of people, movement of a control group in the State.

As a result of the questionnaire application, the direction of care of the project in the Health area and the addition of new sources of information to be incorporated in the georeferenced database were obtained.

Development stage

Through the method of identifying chronic disease patterns using the data mining technique – sorting task[8], it was possible to categorize the chronic diseases: Type 2 DM, Neoplasms, Lung Diseases, Cerebrovascular Disease Hypertension, Obesity, Psychiatric Diseases, and Ischemic Heart.

Thus, a set of records of interest to chronic diseases was selected, without individual identification, analyzing only the information on geographic locations. Information on the health care network (Figure 1) were also implemented; for the process of integration of network information, an ETL tool was used (extraction, transformation, and loading of data) – the Pentaho, which allows one to intuitively define them graphically, thus enabling the documentation of the entire environment[18].
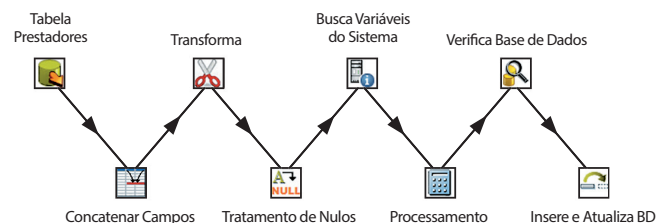
Source: Prepared by the authors.

Figure 1 – Example of a data extraction, transformation, and loading process into the geographical database.

With the information of the health care network layers on the identification of chronic diseases, associated with external information, the layers are integrated into the layers of a geographical database for the construction of a viewing tool.

Based on the formed database, we created through a geographic information system tool – Quantum GIS[17], a free software with general public license (GNU) – several layers for geographic analysis.

Deployment stage

Through the application of categorization rules for identifying beneficiaries with Type 2 DM, a sample of 18,013 individuals was obtained. The epidemiological profile of these individuals is exposed in Table 1. Among the individuals within the sample, 10,495 (58.3%) are female and 7,518 (41.7%) are male. Type 2 DM is predominant in the age group from 60 to 69 years, with 3,910 cases (21.7%), followed by 50 to 59 years, with 3,484 cases (19.3%), 40 to 49 years, with 2,742 cases (15.2%), 30 to 39 years, with 2,553 cases (14.2%) and 70 to 79 years, with 2,494 cases (13.8%). It should be noted these five age groups accumulate 84.3% of the cases. The remaining age groups analyzed (80 years or older and below 30 years of age) represent 15.7% of the cases, of which 8.5% is related

to ages below 30 years, situations of values in the database that recognized possible cases of other types of DM. An important risk factor for diabetes is obesity; in the population studied, 9.4% of the cases (1,702) presented attendances with the ICD of obesity in the health records, which may be E66 – Obesity, E66.0 – Obesity due to excess calories, E66.1 – Drug-induced obesity, E66.2 – Extreme obesity with alveolar hypoventilation, E66.8 – Other obesity, and E66.9 – Obesity, unspecified.

The geographical mapping of the beneficiaries analyzed was based on the State of Paraná, the main area of practice of the health plan operator. Individuals in the sample were prepared according to the city of residence and compared to the total number, with the purpose of knowing in which municipalities the eligible individuals were. Results can be analyzed through Table 2. Twenty-five cities in the State of Paraná concentrate 80.65% of the beneficiaries categorized as Type 2 DM in 2017. In the remainder, 374 cities concentrate 19.35% of the individuals with type 2 DM. In this table, the N of beneficiaries of the HPO decreases because not all of them reside in the State of Paraná.
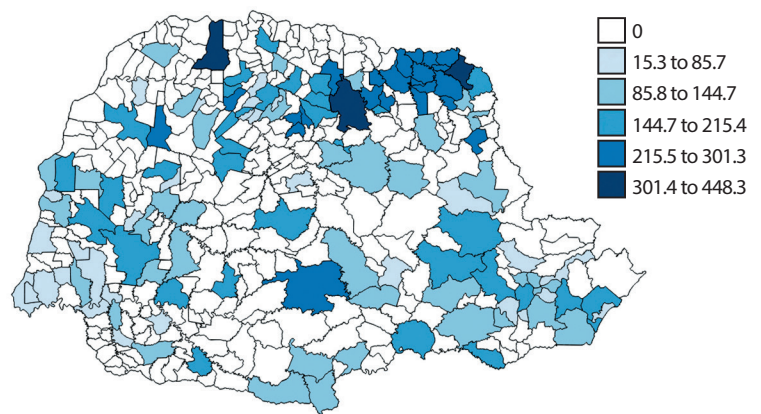
Seeking to meet the study goals and providing different perceptions of visualization, the individuals in the sample were arranged in a georeferenced way, using the software application mentioned. In Figure 1, one can observe that the pioneer North, Central North, and Central South regions are the ones with the highest percentage of individuals within the sample. Georeferenced analysis allows the geospatial visualization of cases, enabling knowing in which regions of the State one must invest in preventive actions and health promotion.

Source: Prepared by the authors.

Figure 2 – Geographic distribution of cases (incidence) of beneficiaries of the health plan operator categorized as Type 2 Diabetes Mellitus, Paraná, Brazil, 2017, according to municipality of residence.

(355.42/1000 beneficiaries), Paranavaí (340.38/1000 beneficiaries) and Guarapuava (301.57/1000 beneficiaries). The research was conducted in the own database, with cases categorized as Type 2 DM. The incidence was calculated based on the number of HPL beneficiaries dwelling in each municipality – the numerator was the Type 2 DM and, as the denominator, the number of beneficiaries in this period, multiplied by 1,000. Data from the municipalities of the state with a sample smaller than 50 were excluded.

Table 2 – Beneficiaries of the health plan operator categorized as prone to Type 2 Diabetes Mellitus, Paraná, Brazil, 2017, according to municipality of dwelling

City	n (N= 15.572)	%	Cumulative %
Curitiba	5850	37.57	37.57
Londrina	1663	10.68	48.25
Maringá	744	4.78	53.02
Cascavel	626	4.02	57.04
Ponta Grossa	346	2.22	59.27
Guarapuava	286	1.84	61.10
Foz do Iguaçu	265	1.70	62.81
São José dos Pinhais	230	1.48	64.28
União da Vitória	213	1.37	65.65
Pinhais	200	1.28	66.93
Bandeirantes	183	1.18	68.11
Paranavaí	177	1.14	69.25
Apucarana	173	1.11	70.36
Jacarezinho	165	1.06	71.42
Paranaguá	153	0.98	72.40
Campo Mourão	148	0.95	73.35
Assaí	147	0.94	74.29
Toledo	140	0.90	75.19
Umuarama	139	0.89	76.09
Cornélio Procópio	137	0.88	76.97
Ubiratã	123	0.79	77.75
Cambará	121	0.78	78.53
Uraí	114	0.73	79.26
Andirá	108	0.69	79.96
Arapongas	108	0.69	80.65
Other municipalities in the state of Paraná	270	19.35	100

Fonte: Elaborado pelos autores.

Table 1 – Epidemiological profile and association with obesity records of the beneficiaries with Type 2 Diabetes Mellitus of the health plan operator, Paraná, Brazil, 2017

Variables	n (N= 18.013)	%
Sex		
Female	10.495	58.3
Male	7.518	41.7
Age Group		
60 to 69 years old	3.910	21.7
50 to 59 years old	3.484	19.3
40 to 49 years old	2.742	15.2
30 to 39 years old	2.553	14.2
70 to 79 years old	2.494	13.8
80 years or older	1.305	7.2
Below 30 years old	1.525	8.5
Associated obesity		
Yes	1.702	9.4
No	16.311	90.6

Source: Prepared by the authors.

Figure 2 shows an example of the application of the visualization model, with the mapping of HPO beneficiaries categorized as Type 2 DM, where the more intense the colors the higher the incidence of the disease per municipality. The rate ranged from 15.38/1000 to 448.37/1000 beneficiaries. The cities with the highest incidence were: Jacarezinho (448.37/1000 beneficiaries), Londrina

DISCUSSION

Information obtained through the identification of standards using the data mining technique and presented geographically potentiates the discovery of new knowledge in the database and, thus, allows dynamic and agile health management actions.

An example of such practice is the geographic mapping shown in Figure 2, which identifies the need for developing actions related to Type 2 DM in the Pioneering North and North of Paraná, mainly due to its high prevalence. In another study, one could note the concern about the high prevalence of hepatitis B cases in the southwestern region of Paraná. As an action resulting from this finding, together with the health area of HPO, we sought to identify if there were beneficiaries resident in that region who had not immunized against Hepatitis B. Through a survey carried out in the database, with subsequent telephonic contact with the beneficiaries identified, it was found that a large part of the beneficiaries residing in that region, about 200, were not immunized against Hepatitis B. Based on this information, the expert on health management via telemonitoring provided guidance to the beneficiaries regarding the importance of immunization, proactively acting towards health education. Many beneficiaries were unaware of the disease scenario in that region and also that the vaccine is available for all age groups, at no cost, in the health facilities of their cities.

The results of applying this model in geographic distribution, using the data mining method, were efficient for the eligibility of beneficiaries who could potentially evolve to chronic diseases. According to the reference[8], 5,953 beneficiaries were indicated as eligible for the diabetic program in 2011, representing 5.7% of the total beneficiaries within the portfolio. This result is compatible with the Surveillance of Risk Factors and Protection for Chronic Diseases by Telephone Inquiry (Vigitel) of 2012, which showed that 5.6% of the adults ($\geq = 18$ years) had a medical diagnosis of Diabetes[25].

Preventively addressing epidemiological alerts and identifying potential risks that may be mitigated by the adoption of proactive actions is a key role for the HPO sustainability, collaborating and acting in partnership with public institutions focused on the care of the population.

Study limitations

This study presented some limitations. In the integration of some public bases listed with the health of managers of the institution, it was observed that some of them had not been updated. Some bases presented data only up to 2012, reflecting the quality of the databases.

Contributions to the fields of nursing and health

The study contributed to the field of nursing and public health, as the use of georeferencing tools is a current necessity of health services. It should be highlight the research valued the use of a free software tool, without costs for purchase and maintenance, an opportunity that enables the access by any health service.

CONCLUSION

The model proposed and described in this study, which is based on a georeferencing tool applied to health, has proved to be efficient and even useful for uncovering patterns at regional levels and, by a cause-and-effect relationship, shall contribute to the formulation and identification of situations involved in the prevention and promotion of health.

The geographical identification of beneficiaries categorized as prone to chronic diseases, using database recognition methodologies and aggregating the service occurrence information, through medical bill records and HPO release requests, will somewhat enable the identification of alert situations, promoting direct actions of prevention and health promotion in advance, which are linked to the main objective of the proposed methodology.

Products derived from this methodology will be useful for the management of HPO services, such as network management, health technology assessment, medical specialty analysis, among others, enabling managers to use agile and timely information for supporting the decision-making process.

In a future research, we also intend to record the alerts and situations identified by the environment, to evaluate the results the tool provided regarding cost optimization for HPO and as a benefit to its customers.

FUNDING

REFERENCES

1. Almeida Filho N, Rouquayrol MZ. Introdução à Epidemiologia. 4th ed. Rio de Janeiro: Guanabara Koogan; 2006.

2. Ashton-Prolla P, Vargas FR. Prevalence and impact of founder mutations in hereditary breast cancer in Latin America. Genetics Molecular Biol[Internet]. 2014[cited 2017 Feb 03];37(1):234-40. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3983579/pdf/gmb-37-234.pdf

3. Banco Mundial. Banco Internacional para a Reconstrução e o Desenvolvimento. Envelhecendo em um Brasil mais velho: implicações do envelhecimento populacional para o crescimento econômico, a redução da pobreza, as finanças públicas e a prestação de serviços. Washington: World Bank; 2011.

4. Instituto de Estudos de Saúde Suplementar (BR). Envelhecimento populacional e os desafios para o sistema de saúde brasileiro. São Paulo: IESS; 2013.

5. Instituto de Estudos de Saúde Suplementar (BR). Variação de Custos Médico-Hospitalares[Internet]. 2016[cited 2017 Dec 12]. Available from:

https://www.iess.org.br/cms/rep/VCMH_set17.pdf Portuguese.

6.   Instituto Paranaense de Desenvolvimento Econômico e Social. Índice nacional de preços ao consumidor (INPC) e índice nacional de preços ao consumidor amplo (IPCA) – Brasil – janeiro 1994 – novembro 2017[Internet]. 2017[cited 2017 Dec 12]. Available from: http://www.ipardes.gov.br/pdf/indices/inpc_ipca.pdf Portuguese.

7.   Carneiro LAF. Planos de saúde: aspectos jurídicos econômicos. Rio de Janeiro: Forense; 2012.

8.   Carvalho DR, Dallagassa MR, Silva SH. Uso de técnicas de mineração de dados para a identificação automática de beneficiários propensos ao diabetes mellitus tipo 2.[Internet]. 2015[cited 2017 Feb 02];20(3):274-96. Available from: http://www.uel.br/revistas/uel/index.php/informacao/article/view/16018/17648 Portuguese.

9.   Conselho Federal de Medicina (CFM). Resolução nº 1.819, de 22 de maio de 2007. Proíbe a colocação do diagnóstico codificado (CID) ou tempo de doença no preenchimento das guias da TISS de consulta e solicitação de exames de seguradoras e operadoras de planos de saúde concomitantemente com a identificação do paciente e dá outras providências. Diário Oficial da União 22 maio 2007; Seção 1.

10.   Bellazzi R. Big Data and Biomedical Informatics: a challenging opportunity. Yearb Med Inform[Internet]. 2014[cited 2017 Feb 05];9(1):8-13. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4287065/pdf/ymi-09-0008.pdf

11.   Fayyad U, Piatetsky-Shapiro G, Smyth P. From data mining to knowledge discovery in databases. AI Mag[Internet]. 1996[cited 2017 Feb 05];17(3):37-54. Available from: https://www.aaai.org/ojs/index.php/aimagazine/article/view/1230/1131

12.   Casters M, Bouman R, Van Dongen J. Pentaho Kettle solutions: building open source ETL solutions with Pentaho Data Integration. Hoboken: John Wiley & Sons; 2010.

13.   Müller EPL, Cubas MR, Bastos LC. Geoprocessing of data as a management tool in a family health unit. Rev Bras Enferm [Internet]. 2010[cited 2017 Feb 05];63(6):978-82. Available from: http://www.scielo.br/pdf/reben/v63n6/17.pdf

14.   Concaro S, Sacchi L, Cerra C, Bellazzi R. Mining administrative and clinical diabetes data with temporal association rules. Stud Health Technol Informatics[Internet]. 2009[cited 2017 Feb 05];150:574–8. Available from: https://www.ncbi.nlm.nih.gov/pubmed/19745376

15.   Bellazzi R, Sacchi L, Concaro S. Methods and Tools for Mining Multivariate Temporal Data in Clinical and Biomedical Applications, 31st Annual International Conference of the IEEE EMBS. Minnesota 2009:5629-32.

16.   Ministério da Saúde (BR). Vigilância de fatores de risco e proteção para doenças crônicas por inquérito telefônico: estimativas sobre frequência e distribuição sócio demográfica de fatores de risco e proteção para doenças crônicas nas capitais dos 26 estados brasileiros e no distrito federal em 2016. Brasília: Ministério da Saúde; 2017.

17.   QGIS. A Free and Open Source Geographic Information System [Internet]. 2016[cited 2016 Jun 05] Available from: https://www.qgis.org/en/site/index.html

18.   Pentaho. Data Integration, Business Analytics and Big Data[Internet]. c2016[cited 2016 Jun 05]. Available from: https://www.hitachivantara.com/en-us/products/big-data-integration-analytics.html

19.   Agência Nacional de Saúde Suplementar (ANS) [Internet]. Rio de Janeiro: Agência Nacional de Saúde Suplementar (ANS), c2018 [cited 2019 Mar 17]. Available from: http://www.ans.gov.br

20.   Ministério da Saúde (BR) [Internet]. c2013-2018[cited 2016 Sep 09]. Available from: http://www.portalms.saude.gov.br/

21.   Secretaria da Saúde (PR) [Internet]. Curitiba (PR): Secretaria da Saúde; c2019 [cited 2019 Mar 17]. Available from: http://www.saude.pr.gov.br

22.   Secretaria Municipal da Saúde (PR) [Internet]. Curitiba (PR): Secretaria Municipal da Saúde; c2019 [cited 2019 Mar 17]. Available from: http://www.saude.curitiba.pr.gov.br

23.   Ministério da Saúde (BR). Departamento de Informática do SUS [Internet]. 2016[cited 2016 Sep 09]. Available from: http://datasus.saude.gov.br/

24.   SIMPAR: Sistema Meteorológico do Paraná [Internet]. Curitiba (PR): SIMPAR; c2019 [cited 2019 Mar 17]. Available from: http://www.simepar.br/prognozweb/simepar/home

25.   Ministério da Saúde (BR). Vigilância de fatores de risco e proteção para doenças crônicas por inquérito telefônico: Estimativas sobre frequência e distribuição sócio demográfica de fatores de risco e proteção para doenças crônicas nas capitais dos 26 estados brasileiros e no distrito federal em 2012. Brasília: Ministério da Saúde; 2013.