Educación Médica 27 (2026) 101107

Educación Médica

www.elsevier.es/edumed

REVIEW

Proposal for a framework for the use of secondary data in health education

Francisco Winter dos Santos Figueiredo[a,b,*], **Bhárbara Karolline Rodrigues Silva**[b], **Fernando Rodrigues Peixoto Quaresma**[a] and **Erika da Silva Maciel**[a]

[a] *Laboratory of Innovation and Technology in Teaching (LITE), Federal University of Tocantins, Palmas, Tocantins, Brazil*
[b] *Inspectto Institute of Teaching, Research and Innovation, Palmas, Tocantins, Brazil*

Received 23 October 2024; accepted 26 August 2025
Available online xxxx

KEYWORDS
Active methodologies;
Health teaching;
Innovation

Abstract In health education, the use of real data to understand the health situation of the population enables the transfer of knowledge and reveals the local reality and the spread of diseases. The aim of this study was to describe the major sources of secondary data in healthcare and develop a framework for using these data in teaching. A narrative literature search was conducted in PubMed, Scopus and the Virtual Health Library using descriptors related to teaching, secondary data, and health in Portuguese, English and Spanish. In addition, through the meeting of experts and based on the articles found in the review, a unified framework of these data in teaching is created. A total of 79 studies were found, of which 14 were included after meeting the eligibility criteria. The main databases found were from the Ministry of Education, the SUS Department of Information Technology, and other sources (Institute of Health Metrics and Evaluation and Institute of Applied Economic Research) which were included after a meeting with experts. The developed framework consists of three phases and 11 elements, including planning, searching, and extracting data, adapting for teaching, and incorporating data for health teaching. There are a few data sources that can be used for health education, but those proposed are used in this framework, and it is possible to transform the data into knowledge that can be transferred to teaching, considering teachers' knowledge of indicators. Technologies and innovations are considered in teaching.

PALABRAS CLAVE
Metodologías activas;
Enseñanza de la salud;
Innovación

Propuesta de un marco para el uso de datos secundarios en la educación para la salud

Resumen En educación para la salud, el uso de datos reales para conocer la situación de salud de la población permite la transferencia de conocimientos y revela la realidad local y la propagación de enfermedades. El objetivo de este estudio fue describir las principales fuentes

* Corresponding author.
E-mail address: francisco.figueiredo@uft.edu.br (F.W. dos Santos Figueiredo).

https://doi.org/10.1016/j.edumed.2025.101107

de datos secundarios en atención sanitaria y desarrollar un marco para el uso de estos datos en la enseñanza. Se realizó una búsqueda bibliográfica narrativa en PubMed, Scopus y la Biblioteca Virtual en Salud, utilizando descriptores relacionados con la enseñanza, los datos secundarios y la salud en portugués, inglés y español. Además, a través de la reunión de expertos y a partir de los artículos encontrados en la revisión, se crea un marco unificado de estos datos en la enseñanza. Se encontraron 79 estudios, de los cuales 14 se incluyeron después de cumplir con los criterios de elegibilidad. Las principales bases de datos encontradas fueron del Ministerio de Educación, del Departamento de Tecnología de la Información del SUS, y se incluyeron otras fuentes (Instituto de Métricas y Evaluación de la Salud e Instituto de Investigación Económica Aplicada) después de una reunión con expertos. El marco desarrollado consta de tres fases y 11 elementos, que incluyen la planificación, la búsqueda y extracción de datos, la adaptación para la enseñanza y la incorporación de datos para la enseñanza de la salud. Son pocas las fuentes de datos que pueden ser utilizadas para la educación en salud, pero en este marco se utilizan las propuestas, es posible transformar los datos en conocimiento que pueda ser transferido a la enseñanza, considerando el conocimiento de los docentes sobre los indicadores, las tecnologías y las innovaciones que se consideran en la enseñanza.

Introduction

Secondary data are data available in aggregate or individualized form that does not allow the identification of the individuals from whom the data were collected. These data are used to compose repositories or databases and can be made publicly available.[1] In Brazil, the National Health Council, through resolution No. 510 of 2016, proposes that this type of data for carrying out scientific research should not be submitted for analysis by the Research Ethics Committee system, or even by the National Council of Ethics in Research.[2]

However, despite the advancement in the use of secondary data for the production of scientific evidence and the advancement of knowledge, there is still a long way to go to integrate the use of these data for professional training in the health area, which can be seen as a barrier that can limit clinical and practical decision-making in the health area and when reduced, it can provide professional training with knowledge and practices based on the local reality.

When training health professionals (doctors, nurses, physiotherapists, among others), it is necessary to have knowledge of the behavior of diseases, their epidemiology and their appearance in a given place, in order to train highly qualified professionals with a critical understanding of the reality that surrounds you. Although these secondary data are constantly used for research purposes, they appear to be rarely used for teaching, although the establishment of repositories based on primary data for knowledge dissemination is promoted primarily by high-ranking scientific journals and other institutions with global visibility.[3]

Frameworks are considered instruments that allow the optimization and standardization of processes with an emphasis on better performance, commonly used in the area of software development and engineering, but which have also been used in other areas. In healthcare, they can be found for the development and implementation of technology platforms,[4] implementation of processes in information technology related to the collection and measurement of medical data,[5] and guidelines for living practices in the health sector.[6]

Given the gap in tools for standardization and guidance on the use of secondary data for health education and the need to critically evaluate data sources, it is necessary to understand what is needed to evaluate the use of secondary data on the theoretical-practical basis classes and deliver dynamics for interpreting the reality of a population's health. The aim of the present study was therefore to describe the use of secondary data for health education and to develop a framework for their use.

Method

This study was developed in two stages, with the narrative review being the first stage to identify the sources of secondary data with a consultation with experts who are also the authors of this study to validate the data sources found and add more sources based on their experiences, and; the development of a guideline for the use of secondary data in health education, based on the studies carried out by De Sá Leite[7] and Razzaghi.[8]

Search

In the narrative review, searches were carried out in the PubMed, Scopus and Virtual Health Library (VHL) databases, as they are the main databases in the health area. Searches in the references of the investigated studies and in Google Scholar were also carried out to try to reach studies that were not found using the descriptors in the search strategy.

The search strategy was developed with health sciences descriptors (Decs) and Medical Subject Headings (MeSH). The following keywords were used to search the titles, concatenated by the Boolean descriptors OR and AND,

according to each database studied. For teaching, the keywords "teaching", "education", "educación" were used; for data sources, we used "data source", "database", "repository", "healthcare database" and "Fuentes de Información".

Eligibility criteria

Studies published in journals with a peer review process, which used at least one source of secondary health data or which analyzed secondary data related to health education, published in full text, in the last 20 years were included.

Studies that used only primary data, that used secondary data sources that were not public, that did not talk about health-related topics, or that it was not possible to identify the source of the data were excluded.

Extraction

The data were extracted by two researchers independently and managed in the Zotero software. Data related to study characteristics (title, summary), data sources, and aspects to be considered to assess the reliability and validity of the data used by original research were extracted.

Data analysis

Two researchers independently evaluated the sources of data cited in the studies and other sources of data found in the references of these studies, regarding access, data availability and other characteristics as proposed by other guidelines. Based on this information, two researchers evaluated the information on similar characteristics of the aspects to be investigated and proposed the domains and stages of the framework.

Results

In the initial search, 79 studies were found, 6 in PUBMED, 30 in SCOPUS and 43 in the VHL. A total of 17 duplicate references were removed, and 61 studies were selected for title and abstract readings. Of the studies found, 14 were included after reading the titles and full text (Fig. 1).

In the included studies, several data sources were found, of which the most frequently used were that of the National Institute of Educational Studies and Research – INEP (42.9% of the included studies), followed by the database of the Ministry of Education – MEC (four studies, 28.6%) and from the SUS IT Department – DATASUS (4 studies, 28.6%) (Table 1).

Although not found in the studies included in the review and based on the previous experience of researchers specializing in secondary data sources and searches of article references, we decided to include the Global Burden of Disease (GBD) databases maintained by the Institute of Health Metrics and Evaluation (https://www.healthdata.org/research-analysis/gbd), as it is one of the main sources of data on the health situation of populations recognized globally, the Institute of Applied Economic Research (IPEA), maintained by the federal government of Brazil (https://www.ipea.gov.br/portal/dados), from the Brazilian Open Data Portal, as it is a source that provides access to different

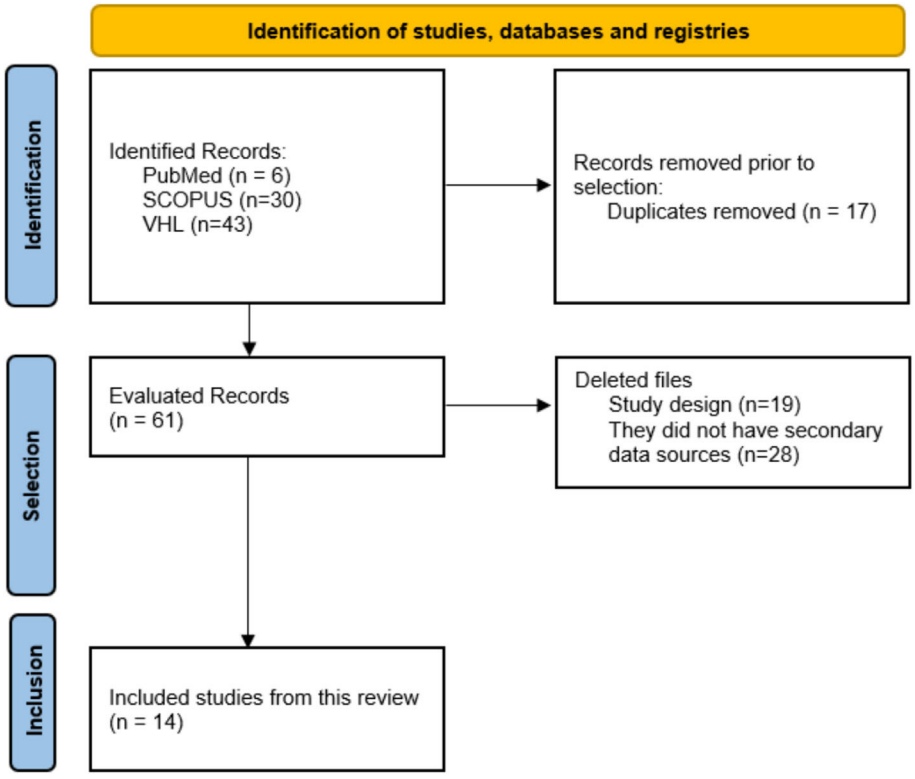

Figure 1 Flowchart of the review steps.

Table 1 Main data sources for secondary data in health education.

Data source	Link
National Institute of Educational Studies and Research (INEP)	https://www.gov.br/inep/pt-br
Ministry of Education (MEC)	http://portal.mec.gov.br/
Department of Health Informatics from Unic Health System (DATASUS)	https://datasus.saude.gov.br/
Brazilian Institute of Geography and Statistics (IBGE)	https://www.ibge.gov.br/estatisticas/sociais/populacao/9127-pesquisa-nacional-por-amostra-de-domicilios.html
Health council monitoring system	https://conselho.saude.gov.br/web_siacs/index.html
Global Burden of Disease (GBD)	https://www.healthdata.org/research-analysis/gbd
Institute of Applied Economic Research (IPEA)	https://www.ipea.gov.br/portal/dados
Brazilian open data portal	https://dados.gov.br/
Health System Performance Assessment Program – PROADESS/FIOCRUZ	https://www.proadess.icict.fiocruz.br

Brazilian data and from the Health System Performance Assessment Program – PROADESS (https://www.proadess.icict.fiocruz.br/) which is maintained by the Oswaldo Cruz Foundation (FIOCRUZ), as it compiles data on the health system and is an easy-to-view platform.

Development of framework

To develop the framework, we used guidelines on the use of secondary data in research, on how to develop guidelines for reporting research and on the steps for developing educational interventions in health. Areco .[3] consider understanding the problem, resource planning, understanding, preparation, validation and distribution of data as the necessary steps for developing teaching tools.[7] Furthermore, we adapted the guidelines to report guidelines for carrying out research proposed by Moher,[9] which details the steps necessary for good presentation of guidelines and the 17 items of the checklist proposed by Phillips[10] to report educational interventions and teaching practice (Table 2).

The developed framework consists of a step-by-step guide for health professionals regarding the use of secondary data obtained from reliable sources, which have been the subject of study in scientific publications. There are three steps and 10 items that detail the actions that must be followed for the good use of secondary data obtained from secondary data sources by health professionals regarding health teaching, which are planning (4 items), data search (3 items) and preparation for presentation (3 items) (Table 3).

Discussion

Our study proposes something challenging, which is the use of technology constantly used for research in teaching, considering the tools and data sources available for this purpose and demonstrating a long way to go regarding the use of secondary data obtained from secondary health education for data estimations to be used in the classroom.

The objective of this study was not to demonstrate the steps for developing the database, but to provide a tool that serves as a guideline so that health teachers can make use of these technologies and improve the teaching-learning process in the classroom, even enabling the use of active methods such as, for example, the flipped classroom and the problem-based learning.[11]

The motivation for developing this guideline is the decentralization of knowledge[12] in a safe way so that health professionals, teachers and other interested parties can have knowledge about the health conditions of populations at the most diverse geographic levels in Brazil. It is

Table 2 Studies included for the development of the framework.

Author	Aim	Study design	Description
Areco[3]	Describe steps for evaluating secondary data sources	Qualitative study	The main steps are understanding the problem, resource planning, understanding, preparation, validation and distribution of data
From Sá Leite[7]	Construction and validation of an instrument for evaluating educational content	Methodological study	The instrument developed showed good reliability, and it was suggested that it can be used by health professionals to create educational content
Razzaghi[8]	Review the literature on data quality	Literature review	Assessment of data quality principles provides a framework for systematic data quality assessment
Phillips [10]	Develop a checklist on how to report educational interventions	Systematic review and validation by experts	The checklist showed good consistency and internal validation

Table 3 Steps and items for using secondary data in health education.

Stage	Description
1. Planning	1. Define the theme and objective of the class.
	2. Find keywords (mesh or Descs) about the outcome to be studied.
	3. Establish the temporality and location of the data that will be collected.
	4. Define which question will be answered with the data.
2. Search	5. Search health databases (PubMed, VHL, SciELO, among others) for articles with this outcome and use terms associated with keywords related to databases ("database", "secondary data").
	6. Search secondary data sources, preferably used in scientific health research, such as those found in the narrative review of the present study, such as DATASUS or GBD.
	7. Extract the data to answer the question specified in item 4.
3. Presentation preparation	8. Transform, when necessary, the data to prepare the presentation.
	9. Adapt search results to class objectives.
	10. Assess the need for adaptations for class inclusion, such as type of fonts, letters, etc. (tools such as canva, figma, among others) that allow adaptation.

necessary to recognize the limitations of secondary data sources, making it necessary to carry out more specific studies on the potential and communication interfaces that can be used in classrooms.[13]

The framework was developed based on these studies and is composed of three stages and 10 items, divided between these stages. It is worth highlighting that the use of these scales for health teaching must consider teachers' knowledge about the indicators, about the use of technologies other than conventional ones in the classroom, and a lesson plan to be followed. Therefore, it is necessary to plan, search for data in databases and adapt the extracted data for use in the classroom according to the objectives of the class.

Secondary health databases are constantly used to carry out research and decision-making because they allow a high number of participants, a well-designed method and the use of more robust instruments.[14] Furthermore, the use of secondary data sources in scientific research allows responding to various research objectives, ranging from the evaluation of public policies to the effectiveness of therapeutic strategies.[3] Unlike studies with primary data, the use of these data does not require ethical evaluation as it is not possible to identify the participants; they are public access databases and allow, in a less costly way, the construction of high scientific quality research.[2]

By identifying, through narrative review, the main databases that are sources of secondary health data related to teaching, and by bringing together the experiences of specialists in secondary data, we believe this is a pioneering study that could serve as a basis for other studies and for the development of innovative teaching tools. Innovation in teaching is essential to keep up with the advancement of science.[11] It is necessary to encourage students and researchers about the importance of tools to provide this innovation process,[15] given that many innovations can be optimized with the use of frameworks and guidelines that allow the distribution of knowledge about technologies.[16]

Teaching about health constantly involves presenting data about the situation of the population in which it is being analyzed,[15] with scientific articles being good sources of data for this purpose. However, although we encourage this use, teachers are limited to the data explored by these studies when they could go beyond the data made available by directly consulting the sources of data used in research. This process of going beyond constantly occurs when conducting research, and therefore, secondary data are widely used for this purpose and are little used for other objectives such as teaching and extension.

Platforms like Institute of Health Metrics and Evaluation, which provides data from the Global Burden of Disease study, allows you to dynamically and interactively view various visualizations of health data, both with regard to the occurrence of diseases and risk factors and causes of injuries, the use of which in the literature points to both for research and as a tool to help data-driven policy decision makers,[17] highlighting the quality of these data for use in health education.

In Brazil, despite the open data policy, the number of visualization platforms that allow teachers to interact with data for teaching purposes and mainly in the health area is still low, with initiatives from some departments of the Federal government through visualization of data from the National Registry of Health Establishments (CNES) in an interactive way, available on the agency's own page,[18] increasingly necessary for the transfer of knowledge in the health area.

Characteristics such as coverage, accessibility, effectiveness and data quality are used to evaluate the quality of routine databases that can be used for research,[8] while the form of data collection, presence of a variable dictionary, and other information about the creation and communication of these data – among others – are necessary items for using secondary data sources in research.

Despite being promising and an interesting tool to be used by teachers, lecturers, and any of the actors in health education, a limitation is recognized that this proposed guideline needs to be validated by teaching professionals in other study designs, such as experimental or quasi-experimental studies and that assess their viability. Furthermore, other study designs for validation with other methods and experts from other areas of science could also be interesting to identify gaps that could not be seen by our specialist researchers.

Conclusion

The framework was developed based on a narrative review and expert opinion and is composed of three stages and 11 items, divided between these stages, making it a useful instrument for teachers for teaching health. It is worth highlighting the use of these scales for health education, and teachers must consider the knowledge of indicators, the use of technologies other than conventional ones in the classroom, and a lesson plan to be followed. Thus, we conclude that it is necessary to plan, search for data in databases and adapt the extracted data for use in the classroom according to the objectives of the class.

Funding

Francisco Winter dos Santos Figueiredo thanks for Programa de Desenvolvimento da Pós-Graduação (PDPG) Pós-Doutorado Estratégico CAPEs, number 88887.801526/2023-00.

Bhárbara Karolline Rodrigues Silva, thanks for Programa de Desenvolvimento da Pós-Graduação (PDPG) Pós-Doutorado Estratégico- CAPEs, number 88887.973582/2024-00.

Conflict of interest

The authors declare that they have no conflicts of interest.

References

1. Boslaugh S. Secondary data sources for public health. Cambridge University Press & Assessment; 2007. doi:10.1017/cbo9780511618802.
2. Guerriero ICZ, de Souza Minayo MC. The approval of CNS Resolution no. 510/2016 is an advance for Brazilian science. Health Soc. 2019;28:299–310.
3. Areco KCN, et al. Operational challenges in the use of structured secondary data for health research. Front Public Health. 2021;9:642163.
4. Herman H, Grobbelaar SS, Grobbelaar SS, Pistorius CWI. The design and development of technology platforms in a developing country healthcare context from an ecosystem perspective. BMC Med Inform Decis Mak. 2020;20:1–24.
5. Rippen HE, et al. Organizational framework for health information technology. Int J Med Inform. 2013;82.
6. El Mikati IK, et al. A framework for the development of living practice guidelines in health care. Ann Internal Med. 2022;175:1154–60. doi:10.7326/m22-0514.
7. de Sá Leite S, et al. Construction and validation of an educational content validation instrument in health. Brazil Nurs Mag. 2018;71:1635–41.
8. Razzaghi H, Greenberg J, Bailey LC. Developing a systematic approach to assessing data quality in secondary use of clinical data based on intended use. Learn Health Syst. 2022;6:e10264.
9. Moher D, Moher D, Schulz KF, Simera I, Altman DG. Guidance for developers of health research reporting guidelines. PLOS Med. 2010;7:e1000217.
10. Phillips AC, et al. Development and validation of the guideline for reporting evidence-based practice educational interventions and teaching (GREET). BMC Med Educ. 2016;16:237.
11. Sanglard L, et al. Active teaching methodologies in health education. RGO – Gaúcha Dent Mag. 2022;70:1–10. doi:10.1590/1981-86372022005020220037.
12. da Silva FCC. Open data visualization in the public sector. Ibero-Am J Inform Sci. 2018;11:523–40.
13. Casanovas MM, Ruiz-Munzón N, Ruiz-Munzón N, Buil-Fabregá M. Higher education: the best practices for fostering competences for sustainable development through the use of active learning methodologies. Int J Sustain High Educ. 2021;23:703–27. doi:10.1108/ijshe-03-2021-0082.
14. Wang SV, et al. Reporting to improve reproducibility and facilitate validity assessment for healthcare database studies V1.0. Pharmacoepidemiol Drug Saf. 2017;26:1018–32.
15. Lycan T, et al. An evidence-based medicine approach for case presentations by trainees. Clin Teach. 2019;16:575–9.
16. Lacasse M, et al. Theories and models in health sciences education – a literature review. Can J Scholars Teach Learn. 2019;10:1–29.
17. Lundkvist A, et al. Policy-makers' views on translating burden of disease estimates in health policies: bridging the gap through data visualization. Archiv Public Health. 2021;79:1–17.
18. Brazil, Ministry of Health. Elastic CNES. National Registry of Health Establishments; 2023. https://elasticnes.saude.gov.br/.