Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 196 (2022) 439–444

Procedia Computer Science

www.elsevier.com/locate/procedia

CENTERIS - International Conference on ENTERprise Information Systems / ProjMAN - International Conference on Project MANagement / HCist - International Conference on Health and Social Care Information Systems and Technologies 2021

The minimum dataset for rare diseases in Brazil: a systematic review protocol

Filipe Andrade Bernardi[a,b]*, Diego Bettiol Yamada[b], Bibiana Mello de Oliveira[c], Vinicius Costa Lima[a,b], Têmis Maria Félix[d], Domingos Alves[b]

[a]*Bioengineering Postgraduate Program, University of São Paulo, São Carlos, Brazil*
[b]*Ribeirao Preto Medical School, University of Sao Paulo, Ribeirao Preto, Brazil*
[c]*Postgraduate Program in Genetics and Molecular Biology, Federal University of Rio Grande do Sul, Porto Alegre, Brazil*
[d]*Medical Genetics Service, Porto Alegre Clinics Hospital, Porto Alegre, Brazil*

Abstract

The Minimum Data Set (MDS) can be used for subsidiarity decision-making and health planning. Besides, this strategy allows to identify obligatory points that must be adjusted to achieve sustainable management in the planning and development of relevant Health Information Systems for public health. Specifically, in the context of rare diseases, the MDS strategy can be very valuable. This systematic review will focus on research using MDS for rare diseases in several databases. We seek to answer the question: "What is the minimum data set used in registries for rare diseases?" Some outcomes of interests specific for MDS will involve information about epidemiology, clinical procedures, and therapeutic resources among other features. We hope that by standardizing data through a careful analysis of evidence from different sources of a common format, with shared specifications and structures, we can help in the methodological transparency and reproducibility of results in the context of rare disease research.

Keywords: Rare diseases; Minimum data set; Public Health; Digital Health; Review protocol;

* Filipe Andrade Bernardi. Tel.: +55 16 997880795.
E-mail address: filipepaulista12@usp.br

10.1016/j.procs.2021.12.034

1. Introduction

The Minimum Data Set (MDS) is a public health strategy adopted by many countries to standardize the essential data for its health systems. The MDS can be used for subsidiarity decision-making and health planning. Besides, this strategy allows us to identify obligatory points that must be adjusted to achieve sustainable management in the planning and development of relevant Health Information Systems (HIS) for public health, constituting a fundamental information base for monitoring resources and generating indicators [1].

The use of consistent and validated MDS can meet the growing demands of health management, such as to reduce the fragmentation of information from clinical and administrative HIS, improve the formulation and monitoring of health policies, generate national health statistics, and promote efficient allocation of resources and financing in health, as well as the use of new models for performance and satisfaction measurement [2].

According to the World Health Organization (WHO), one of the pillars for the development of HIS is the functionality to generate indicators that can summarize information relevant to a given health phenomenon and monitor the quality of care, services, and treatments offered to the population [3]. We also know that the generation of quality indicators requires a set of raw data as a basis. Such data is interpreted, organized, and transformed into information. Then, this information can be processed and analyzed, generating knowledge and intelligence in health [4].

Specifically, in the context of Rare Diseases (RD), the MDS strategy can be very valuable. Studies show that, in this area, a patient's diagnostic process is highly costly and complex, and requires the use of standardized clinical information for greater accuracy. A RD is a medical condition of low prevalence compared to the prevalence of other diseases in one location. Together, RD can represent 10% of all diseases in the population, so they have a significant impact on health systems, and the information needed for evidence-based decision-making should be available for health professionals and managers [5;6].

It is estimated that the number of patients affected by a RD is approximately 30 million in Europe and 25 million in North America [7;8]. As a result, many countries have initiated national plans to promote care, research, and technology in RD. The establishment of a consistent and validated MDS is essential and has a great potential to assist public policies in attending to the needs and demands of the population. The use of this information also enables health managers to improve the services provided [9].

Therefore, the RD domain needs increasingly contextualized, accurate, and standardized information to reduce the fragmentation of the high volume of data, creating a solid base of information pertinent to diagnoses, treatments, and processes [10]. Thus, the objective of this article is to identify the minimum data set for RD used in different countries through a systematic review methodology. It is expected to make this knowledge available to improve the Brazilian MDS for RD.

The methodology presented in this paper is based on studies from the literature, which provide guidelines for conducting reviews that can be applied to several research domains in a nonspecific way. Therefore, in the next section, we will present the related work. In the third section, the methodological approach and key concepts of review will be described. Finally, in the fourth and fifth sections, expected contributions will be highlighted, as well as possible limitations, conclusions and future work.

2. Background

In the Brazilian Public Health System, we observed the existence of a long therapeutic itinerary of a diagnosis of a RD. It may involve directing the patient to different healthcare units of varied specialties until the final result, in addition to the difficulty of access to specialists, exams and specific therapies [11]. Thus, the MDS was instituted in this country by the Decree of November 29th 2017, presented as an essential component for the surveillance of epidemiological issues and the provision of health services [12]. The MDS in RD also allows detecting trends that manifest each of the indicators recommended and evaluated in the Brazilian Policy for Comprehensive Care for People with Rare Diseases (BPCCPRD) of the Ministry of Health of Brazil [13].

In France, national health managers have identified the need for better coordination of services and care related to RD in the country. In this case, a minimum set of clinical data was proposed to standardize the sharing of information on RD patients between centers specialized in this area, bringing benefits to users of the health system in that country. The French MDS was accepted by national consensus, promoted better coordination and management of health

services, and facilitated the dissemination of information for academic studies, bringing an evolution to the area of RD [1].

In the United Kingdom, scientists have developed a web-based data registration method for rare pediatric lung diseases, since they identified that patients received less than ideal care due to a lack of understanding of pathophysiology of the disorder. Then, they demonstrated that the establishment of a standardized data set for the records of these patients was successful, bringing better clinical and administrative outcomes. The need for other European countries to collaborate with the establishment of standardized data records was also raised, as a matter of urgency, as a tool to reduce inequality in health care [14].

3. Methods

The systematic review will contemplate the search for research about MDS for RD in several databases. The search strategy aims to find both published and unpublished studies. As recommended by the Joanna Briggs Institute (JBI), a three-step search strategy will be applied. The requirements of data sharing within the RD realm are compelling, i.e., simply data exchange is not enough. Our proposal intends to demonstrate that effective data sharing networks require consensus on data syntax, meaning, and security of information in the Brazilian scenario. In this case, the solution must be able to understand and communicate minimally with other systems and need to be according BPCCPRD and global literature.

The intended review, described on this protocol, is part of an umbrella project that is a design mixed prospective and retrospective observational cohort study to map the scenario of RD in Brazil. The current review protocol is registered in the PROSPERO platform - International Prospective Register for Systematic Reviews (registration number CRD42021221593).

3.1. Stage 1: Identifying the research questions

In this study, we seek to answer the question: "What is the minimum data set used in registries for rare diseases?" through a systematic review. Then, we intend to compare and propose suggestions for the Brazilian MDS in the context of RD, following the WHO premises.

3.1.1. Eligibility criteria

This review will include research (full papers, conference papers) and non-research studies (editorials, narrative reviews) that present methodologies of a fully functional computerized tool, regardless of a graphical user interface, that could help obtain any domain extraction related to RD. Furthermore, any study design (i.e., quantitative, qualitative, or mixed methods) as well as opinion pieces, commentaries, letters, and editorials, will be considered. National plans and policies, industry reports, position papers, or program reports will also be included. Review articles will be excluded, but relevant papers will be used to crosscheck for primary papers. Personal blogs and social media posts will be excluded, as well as only theoretical approaches without a practical implementation, such as models' definitions (operational, predictive) or ontologies. Also, selected works must address RD in humans. Only English and Portuguese language papers, not limited by date, will be included.

3.1.2. Search strategy

A first limited search in PubMed database will be performed followed by an analysis of the text words contained in the title and abstract, and the index terms available at MeSH (Medical Subject Headings), a vocabulary thesaurus used for indexing articles for PubMed [15]. A second search using all identified keywords and index terms will then be undertaken across all included databases. Lastly, the reference list of all identified reports and articles will be searched for additional studies.

Six databases were chosen: PubMed, CINAHL (Cumulative Index to Nursing and Allied Health Literature), Scopus, EBSCO (Health Source - Consumer Edition), MEDLINE and Web of Science. Documents from the World Health Organization (WHO) and governmental websites will also be examined for policies and reports. The search for gray literature studies will include Google Scholar, ProQuest Dissertations and Theses, Open Thesis, BDTD

(Brazilian Digital Library of Theses and Dissertations), and NDLTD (Networked Digital Library of Theses and Dissertations). The search terms will be “rare disease”, “digital health”, “minimum dataset”, “national network” and "national plan". The search strategy will be composed by the combination of the referred terms with the Boolean terms: AND, OR and NOT composing the search formulas in the referred databases and search platforms (Table 1).

Table 1. Search strategy.

Database	Search string
Pubmed, MEDLINE, SCOPUS, Web of Science	("rare disease") AND ("digital health" OR " minimum dataset ") AND ("national network " OR ""national plan")
EBSCO, CINAHL	Doença Rara AND Saúde Digital OR Conjunto Mínimo AND Rede Nacional OR Plano Nacional

3.2. Stage 3: Study Selection

After carrying out searches adopting the terms in the mentioned databases, two investigators will independently screen each retrieved article based on title and abstract for eligibility according to the inclusion and exclusion criteria. Then the full text will be retrieved, and the investigators will independently perform another round of review to determine if these full texts meet the eligibility criteria. Reviewers will not be blinded to the journal’s title, study authors, or associated institutions. Divergences between the two reviewers will be discussed. Should reviewers not reach an accord, a third reviewer will be participating in the process to make a final decision on inclusion or exclusion. Search results and the eligibility screening process of this phase will be reported using the PRISMA flowchart to detail the numbers of each review process [16].

3.3. Stage 4: Charting the data

Data to be extracted will likely include study characteristics such as author(s), publication year, title, country of publication, national context, study objective(s), sample size, data collection method, data analysis method and main findings and strength of evidence recommendation according to the quality assessment tool (Table 2). Thus, this step of charting will be iterative, enabling the reviewers to update the data extraction form.

Table 2. Data extraction strategy.

Domain	Data to be extracted
Summary	Author(s), title, citation, publication type, country of origin, aims/objectives, if it is a fully functional tool and if it contains a graphical user interface.
Question 1: addressing topic	Type of technology and/or methodology used
Question 2: Domain categorization	MDS main applicability (clinical, epidemiological/managerial)
Question 3: outcomes and gaps	Outcomes of interest, limitations

3.4. Stage 5: Collating, summarizing and reporting the results

In the results section, data will be visually represented in tables and diagrams using data extracted from the included papers. According to the type of studies, the review output will be presented in a descriptive format that aligns with the objectives and scope of the review. The narrative synthesis will seek to investigate similarities and differences between studies to explore patterns, themes, and relationships and propose explanations for findings, e.g., how and why certain MDS development or implementation strategies have worked, or not, in the RD field.

3.5. Stage 6: Quality appraisal

The methodological quality of the included systematic reviews will be assessed by two independent reviewers using the multiple systematic reviews assessment tool (AMSTAR 2). Regarding the quality evaluation of the results and the strength of the evidence recommendations found in our review, the GRADE approach (Grading of Recommendations Assessment, Development and Evaluation) will be adopted [17].

4. Results and discussion

Some outcomes of interests specific for MDS will involve information about epidemiology, clinical procedures, and therapeutic resources among other features. Besides that, we hope to find standards in socio-demographic, epidemiological, clinical, and therapeutics data, able to identify and characterize the type of treatments, the existing diagnostic and technological resources within networks. Also, it is expected to map concerns about human resources to support information about what specialists are required to elaborate a network according to the guidelines of WHO and the national plans.

Then, according to our goal in this work, we hope to set forth a standard for the minimum information required to figure out the aspects of RD populations and their possible hidden biases. This finding should assist researchers, policy makers and system engineers to think about the conception and methodologies used for creating clinical decision tools and surveillance systems.

As possible limitations, relevant non-English language publications may be excluded. To mitigate this, we mapped validated MDS models for rare diseases using the main scientific bases of the area, and the main keywords that encompass the concepts of interest, as described in the Methods section. Besides, our domain includes all rare disease networks in healthcare systems.

Subsequently, we will contrast these findings with the model used in the BPCCPRD. Thus, we hope to be able to identify needs, specify improvements and point out suggestions for the Brazilian rare disease MDS, aiming to build an information structure with greater conformity and alignment with international strategies, validated through evidence-based practice. However, it will be done always considering the specific health demands of each region.

Our results can also be used to improve communication between the care units that make up the Brazilian Public Healthcare System. Through a standardized structure of the minimum information that must be collected for patients diagnosed with rare diseases, governmental strategic groups can be notified, enabling the rapid identification of demands and the development of adequate health policies to meet health services and patients' needs, in primary health care units, secondary care units, and highly complex general hospitals.

With a consistent and validated MDS, it is also possible to map the management processes of a health unit or hospital and identify points in the process flow where the minimum necessary data can be obtained. In this way, it is more viable to define clearly and in a transparent and didactic way to employees and patients what information needs to be collected at what point in the process at the institution, and how this relates to the MDS of the Ministry of Health of Brazil to promote the improvement of the services offered in the area of RD in the country.

Indirectly, this step can provide improvements and benefits to health managers, allowing greater efficiency in services, increased quality of service, and better allocation of resources. Promoting direct benefits to society through the relevant data to impact the performance and effectiveness of the public health sector services.

5. Conclusion and future work

The protocol proposed in this work may result in the first systematic review conducted to provide an overview of the MDS of RD. The results will be used to provide a broad understanding of their adoption for better planning and dissemination at the national levels according to each country's domain particularity. In conclusion, we expect that the outcomes will bring evidence about the best practices using the MDS approach for these networks that involve many stakeholders, e.g., hospitals, universities, and primary healthcare units. Furthermore, the information gathered in this study can provide subsidies for better decision-making in the area of RD and also for planning and creating health policies.

Thus, we hope that by standardizing data through a careful analysis of evidence from different sources of a common format, with shared specifications and structures, we can help in the methodological transparency and reproducibility of results in the context of RD research. When discussing the validity and purpose of these datasets, one must distinguish not only between their intended use, such as use in research or improving the quality of services, but also between the diseases included and treatment procedures addressed. Also, it is expected that the mapping of concepts from different sources allows us to understand processes and tools that contribute to a set of characteristics that change over time in the same data environment, and thus it is possible to measure the severity of the impact of their changes.

As future work, we can mention the use of the data models described in this study to serve as a structured informational basis for the development of decision-support information systems and health observatories that are increasingly relevant to the public health, concerning the context of RD.

Acknowledgements

This study was supported by CNPq (#443030/2019-7), TMF is supported by CNPq (#306861/2019-4). This research is under development and is part of the expected results of The National Network of Rare Diseases project. It was approved in Edital nº. 25/2019 from CNPq (‘Brazilian Council for Scientific and Technological Development, in Portuguese) with financial support from the Ministry of Health, in the amount of 3.5 million reais.

References

[1] Choquet, Rémy, et al. (2015) "A methodology for a minimum data set for rare diseases to support national centers of excellence for healthcare and research." *Journal of the American Medical Informatics Association* **22**(**1**): 76-85.

[2] Werley, Hamet H., et al. (1991) "The Nursing Minimum Data Set: abstraction tool for standardized, comparable, essential data." *American journal of public health* **81**(**4**): 421-426.

[3] World Health Organization. (2015) “Global reference list of 100 core health indicators.” [accessed in June 7, 2021]. Available on: http://apps.who.int/iris/bitstream/handle/10665/173589/WHO_HIS_HSI_2015.3_eng.pdf%3Bjsessionid=91D7FB3935866ABCF0BDE11BF52BC1F4?sequence=1

[4] Yamada, Diego Bettiol, et al. (2020) "Ontology-Based Inference for Supporting Clinical Decisions in Mental Health." *International Conference on Computational Science* **2020**: 363-375.

[5] Richter, Trevor, et al. (2015) "Rare disease terminology and definitions—a systematic global review: report of the ISPOR rare disease special interest group." *Value in Health* **18**(**6**): 906-914.

[6] Alves, Domingos, et al. (2021) "Mapping, Infrastructure, and Data Analysis for the Brazilian Network of Rare Diseases: Protocol for the RARASnet Observational Cohort Study." *JMIR Research Protocols* **10**(**1**): e24826.

[7] Donnai, D., et al. (2011) "2011 Report on the State of the Art of Rare Disease Activities in Europe of the European Union Committee of Experts on Rare Diseases-Part I: Overview of Rare Disease Activities in Europe and Key Developments in 2010 (United Kingdom)." *Annual reports on the state of the art of rare diseases activities in Europe*.

[8] US Food and Drug Administration. (2017) "Developing products for rare diseases & conditions." *US FDA website*.

[9] Bodí, María, et al. (2021) "Automatic generation of minimum dataset and quality indicators from data collected routinely by the clinical information system in an intensive care unit." *International Journal of Medical Informatics* **145**: 104327.

[10] Bellgard, Matthew I., Tom Snelling, and James M. McGree. (2019) "RD-RAP: beyond rare disease patient registries, devising a comprehensive data and analytic framework." *Orphanet journal of rare diseases* **14**(**1**) : 1-7.

[11] Iriart, Jorge Alberto Bernstein, et al. (2019) "Da busca pelo diagnóstico às incertezas do tratamento: desafios do cuidado para as doenças genéticas raras no Brasil." *Ciência & Saúde Coletiva* **24**: 3637-3650.

[12] Brasil. (2016) “Portal CMD do Ministério da Saúde" [accessed in June 7, 2021]. Available on: https://conjuntominimo.saude.gov.br/#/ultimas-noticias

[13] Brasil. (2017) “Decreto de 29 de Novembro de 2017” [accessed in June 7, 2021]. Available on: http://www.planalto.gov.br/ccivil_03/_Ato2015-2018/2017/Dsn/Dsn14501.htm

[14] Laverty, A., A. Jaffe, and S. Cunningham. (2008) "Establishment of a web-based registry for rare (orphan) pediatric lung diseases in the United Kingdom: The BPOLD registry." *Pediatric pulmonology* **43**(**5**): 451-456.

[15] Gasper, William, Parvathi Chundi, and Dario Ghersi. (2020) "MeSH Indexing Using the Biomedical Citation Network." *Proceedings of the 11th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics* **2020**: 1-8.

[16] Moher, David, et al. (2010) "Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement." *Int J Surg* **8**(**5**): 336-341.

[17] Brignardello-Petersen, Romina, et al. (2018) "Advances in the GRADE approach to rate the certainty in estimates from a network meta-analysis." *Journal of clinical epidemiology* **93**: 36-44.