# Strategic Assessment of Healthcare Data Management Through SWOT Analysis

**Rasmussen Stu**
*Department of Economics, University of Craiova, Romania*

**Noel Appleby**
*Department of Political Theory, University of Bucharest, Romania*

*The lack of data for many indicators and the existence of significant gaps in the availability and comparability of data is a problem of global interest, felt by authorities and agencies specialized in data collection, analysis and use. Their objective is to provide data and indicators of good quality, which will help to correctly inform the political decision-makers and to solve the inequalities in the field of health. Thus, this paper aims to provide clear and coherent solutions to a number of problems identified in data management in general and in the medical sector in particular. The added value of this paper consists in a case study, which presents an x-ray of the current situation of data quality through the analysis of metadata. The research methodology used in the paper includes a swot analysis on the quality of health data, using metadata from three sources corresponding to distinct levels of aggregation: National (Romania), European and Global.*

**Keywords:** Health, Data Management, Data Quality, Metadata, Real-World Data

## INTRODUCTION

If we are to analyse the global context, for about a fifth of the countries, over half of the indicators lack of primary data or recent underlying data. There are significant gaps regarding the availability and quality of data for the global health monitoring. Many national health information systems require urgent improvements starting with greater investments in human resources, technology and collaborations. By collecting, analysing and using higher quality data it is made an important step towards improving and solving health inequalities. The development of integrated health statistics is very important as they will represent key information for policy makers. Considering the COVID-19 emergency as an example, a functional system must flexible enough to adapt to unexpected situations in order to provide relevant data on time for key decisional factors. (World Health Organization, 2020)



International Journal of Stress Management
2021, Vol. 28, No. 3, 1–11
© 2021 American Psychological Association

Despite substantial progress in recent decades, national statistical systems and the health data generated by them are characterized by significant limitations.

Disaggregated data, which are essential for health inequalities monitoring, are often missing. However, when primary data are not sufficiently comparable, approaches such as mathematical or statistical models are used to produce comparable estimates, based on the use of available primary data. (World Health Organization, 2020)

Household surveys currently remain the main source of disaggregated health data on demographic and socio-economic characteristics. However, the samples often do not have the necessary dimensions to allow the calculation of sub-national statistics beyond the regional level or the measurement of health indicators in disadvantaged or difficult-to-study populations (such as migrants or minorities). (World Health Organization, 2016) In this regard, real-world data (RWD) can be used to generate actionable safety evidence more quickly and has the potential to produce substantial financial savings. (Bartlett et al., 2019) This is why RWDs are now increasingly used for the purpose of regulation or evaluation of medical technology. There is widespread recognition that the use of health data to advance research and development of personalized medicine has the potential to provide significant benefits to patients and health systems. Physicians are increasingly receiving therapies supported by data analysis, which have the potential to replace trial and error with accurate diagnosis and treatment, benefiting both the provider and the beneficiary of health and care services. (European Commission, 2018)

## DATA LIMITATIONS AND RELATED SOLUTIONS

### The problem of data fragmentation

Currently, there is a significant fragmentation of health data available in the European Union (EU), as most health statistics focus on certain diseases or particular care settings (e.g. hospitals) without a person-centered perspective. Patient-centered data moves away from the limitations of the disease in order to capture patient outcomes, thus facilitating a comprehensive approach to individual health care. (European Commission's Directorate General of Health and Food Safety, 2017) For example, we should know after the treatment, whether the patient is personally satisfied, if he possesses a full range of functional abilities, with an unhindered social life and the ability to return to daily activities. Patient- centered data should also cover the entire journey through the health system. In this regard, indicators linking pre-treatment events to subsequent outcomes could help to better understand the effects of medical interventions on

disease evolution and subsequent patient well-being. (European Commission's Directorate General of Health and Food Safety, 2017)

However, a precondition for making patient-centered data a reality is to create a coherent data governance framework, with explicit rules and guarantees of confidentiality. The General Data Protection Regulation (GDPR), applicable from 25th of May 2018, provides a uniform set of data protection rules throughout the EU. The GDPR is directly relevant for the digitization of health and Article 9 defines special provisions applying to personal health data. (European Commission, 2018)

## Lack of data and comparability issues

In order to address and fill data gaps, it is important to understand the state of a country's health information system. Using SCORE (survey, count, optimize, review, enable) for data sets, countries can observe strengths and weaknesses, identifying gaps in national health information systems. The World Health Organization (WHO) works with countries and health agencies to produce global and national estimates for global indicators, helping to ensure that data are comparable and of high quality. To do this, WHO supports countries in collecting, analysing, using and sharing data. (World Health Organization, 2020)

Given the significant discrepancies in data and the lack of timely data for many indicators, it is often necessary to use statistical models to get a picture of the global and regional situation, including comparable statistics for use by country. These estimates differ from nationally reported data, which are often not adjusted or do not relate to the same year. Thus, further efforts should be made to harmonize the data provided globally with the data published by the national statistical authorities. (World Health Organization, 2016)

## Big data and cross-border access

There has been a significant increase in the quantity and quality of health data generated by new devices over the last decade. Improvements in computing capacity and performance, as well as the emergence of other digital technologies that allow high storage capacity and advanced data analysis, make it possible to accelerate research and development of new treatments, as well as the prevention and early diagnosis of disease. Personalized medicine has thus gained widespread recognition as a useful tool for healthcare organizations in different areas of the disease. (European Commission, 2018)

Big data in the health sector include large volumes of various information (biological, clinical, environmental and lifestyle), collected from both individuals and large cohorts, in relation to health and well-being, at one or more moments in time. A major challenge that health authorities, practitioners and scientists are

facing is the ability to bring this wealth of data together and interpret a finding in the context of the individual patient. Connecting these data sources offers the possibility of early warning, detection of threats to infectious diseases, increased surveillance and control of infectious outbreaks and allows a quick and personalized treatment of infected patients. (European Commission, 2018)

The scientific community is more and more involved in a debate on the benefits of creating a sustainable and efficient ecosystem that brings together cross-border access to data sets, the computing power necessary to analyse and process this volume of data and the scientific knowledge and expertise needed to translate big data into real results for health systems across the EU. (European Commission, 2018)

## From survey data to real-world data

Disaggregation of data is a major challenge for many countries, as well as for global and regional monitoring. Disaggregation of data by characteristics such as gender, age, ethnicity, geographical location or income status reveals vital specifications for identifying and tracking disadvantaged populations within countries. Beyond disaggregation, data processing should be an inclusive and transparent process. Recent developments in technology, such as the use of mobile technology to collect data, provide opportunities to improve statistics for participatory accountability and monitoring. (Abualghaib et al., 2019)

Randomized clinical trials (RCTs) are generally considered to be the standard criterion for the generation of clinical evidence. However, patients selected for RCT often have characteristics, experiences, and treatment protocols that differ from those of patients in daily practice and therefore it may not be possible to generalize the information to a wider group of patients. Moreover, due to the short-term follow-up of randomized clinical trials, the long-term health outcome may not always be visible. (European Commission, 2018) RWD's goal is to provide early access to innovative health interventions, while ensuring safety and efficacity. At the same time, RWD’s strength relies in its ability to improve the generalizability of clinical research findings and to contribute to knowledge, allowing the study of larger cohorts and longer follow-up times than those usually used in RCTs. RWD can also inform policy makers to support the option associated with lower morbidity or lower costs when two treatment options are associated with similar survival outcomes. (Booth et al., 2019)

Regarding the ethical limitations and challenges, there are two important issues to consider when using RWD. The first one is concerning the quality of RWD, as well as the scientific methods used. A huge range of variables may lead the users of the data to draw conclusions based on false correlations about the causal nature of the observed associations. The second issue makes reference to the possibility that those collecting or using RWD may be driven by political,

professional, commercial or even personal interests, due to the fact that the standards for RWD quality are just beginning to emerge. (Lipworth, 2019)

## METHODOLOGY

The paper begins with a qualitative research that brings to light issues such as data fragmentation, data gaps, data comparability and the need for disaggregated data. For this series of problems, we identified several corresponding solutions, such as patient-centered data, methods for data sets optimization, improving international collaboration, the use of big data and cross-border access to data sets. To these solutions is added the use of real-world data, which support the obtaining of disaggregated data, thus completing the data obtained through randomized clinical trials.

Regarding the case study, we performed a SWOT analysis for several indicators, using metadata available on the National Institute of Statistics in Romania, on the Eurostat and on the World Health Organization. The analysis is based on five indicators in the field of health: medical staff, hospital beds, self-perceived health, infant mortality rate and health expenditure. These indicators are compared according to several important aspects: definition, methodology (calculation method and unit of measure), periodicity (including the last update), degree of coverage (the level to which data are available), degree of granularity.

The paper aims to capture through this analysis the lack of data, the problems of availability and comparability of indicators from several statistical sources. Finally, several directions for improving the quality and comparability of the data collected are proposed.

## DATA QUALITY ANALYSIS - RESULTS AND DISCUSSIONS

In this case study, we develop a critical analysis of the metadata by using the SWOT analysis, which will reflect the data quality for the main proposed indicators, at several levels of aggregation: national level (for Romania), European level and global level. To achieve this, we chose a series of indicators from the health sector: health personnel, hospital beds, self-perceived health, infant mortality rate and health expenditure. These five indicators are compared from the perspective of periodicity, the degree of coverage, the granularity, the methodology, as well as the way in which they are defined by the three sources.

### Introduction into data quality and metadata

By analysing the quality of the data it is scientifically determined whether the data obtained from a certain environment are suitable in terms of their type, as well as in terms of the quantity and quality necessary for their use for the purpose for which they were obtained. (Ciora et al., 2003) The Quality Guide for Romanian Official Statistics offers a perspective on how to ensure the quality of research and the statistical process. The National Institute of Statistics defines the quality of results in accordance with the definition given by the European Statistical System (ESS). In broader terms, the quality of National Institute of Statistics results is assessed in terms of "appropriate for use". Specifically, the quality of the result is measured in terms of six components of quality: relevance, accuracy, opportunity and punctuality, clarity and accessibility, coherence and comparability. (Popa et al., 2005)

Statistical metadata represent the information needed to produce and use statistical data. In other words, metadata are data about statistical data. Their role is to help the user to quickly find the information they need and interpret it correctly. Their use facilitates several important actions such as access, exchange, comparison and understanding of statistical data. Within the National Institute of Statistics, the user can access the methodological documents or the instructions for carrying out a certain statistical activity, as well as the statistical forms used. Also developed within the INS, the TEMPO Online database offers the possibility to access a wide range of information provided by the Institute, as well as access to tables with detailed statistical information. (INS, 2020)

Within Eurostat, data on health statistics come from administrative data and population surveys. Administrative data provide a complete and objective picture of topics and usually allow for a regional breakdown. Survey data, which are more subjective in nature, complement administrative data for measuring socio-economic disparities in the use of and access to health care services. Because of the fact that healthcare data are based on sources such as administrative data within states, they reflect the country's specific healthcare organization and may not be completely comparable, with differences in time series coverage or geographical coverage and therefore, the validity and the reliability of data may vary. Due to their heterogeneity, data are collected, stored and disseminated through different tables. (European Commission, 2020)

Moving to global level, we notice that on the WHO website, each entry in the metadata view contains details about the different dimensions that are used to index and describe the information contained in the Global Health Observatory.

Monitoring the health situation, trends, progress and performance of health systems requires data from several sources on a wide variety of health topics. WHO supports Member States by strengthening their capacity to collect, compile, manage, analyse and use health data derived mainly from population-based sources (household surveys, civilian vital registration systems)

and institutional sources (administrative and operational activities of institutions, such as health facilities). (World Health Organization, 2020)

## SWOT analysis of the metadata

SWOT analysis is a general technique for evaluating any public or private organization, helping decision makers to focus on key issues. Carrying out a SWOT analysis requires the generation as well as the recording of the strengths, weaknesses, opportunities and threats related to the organization. In the present study, the SWOT analysis aims to identify the main strengths of the data, the weaknesses that prevent an international alignment, the long-term data processing opportunities, as well as the threats related to data gaps which prevent the comparability of indicators. Among the aspects analysed for each indicator, the definition provided by each institution (INS/ Eurostat/ WHO), the methodology for calculating or estimating the indicators and the degree of disaggregation or granularity were classified as strengths or weaknesses, and the degree of coverage indicator and periodicity of data collection were included in opportunities or threats.

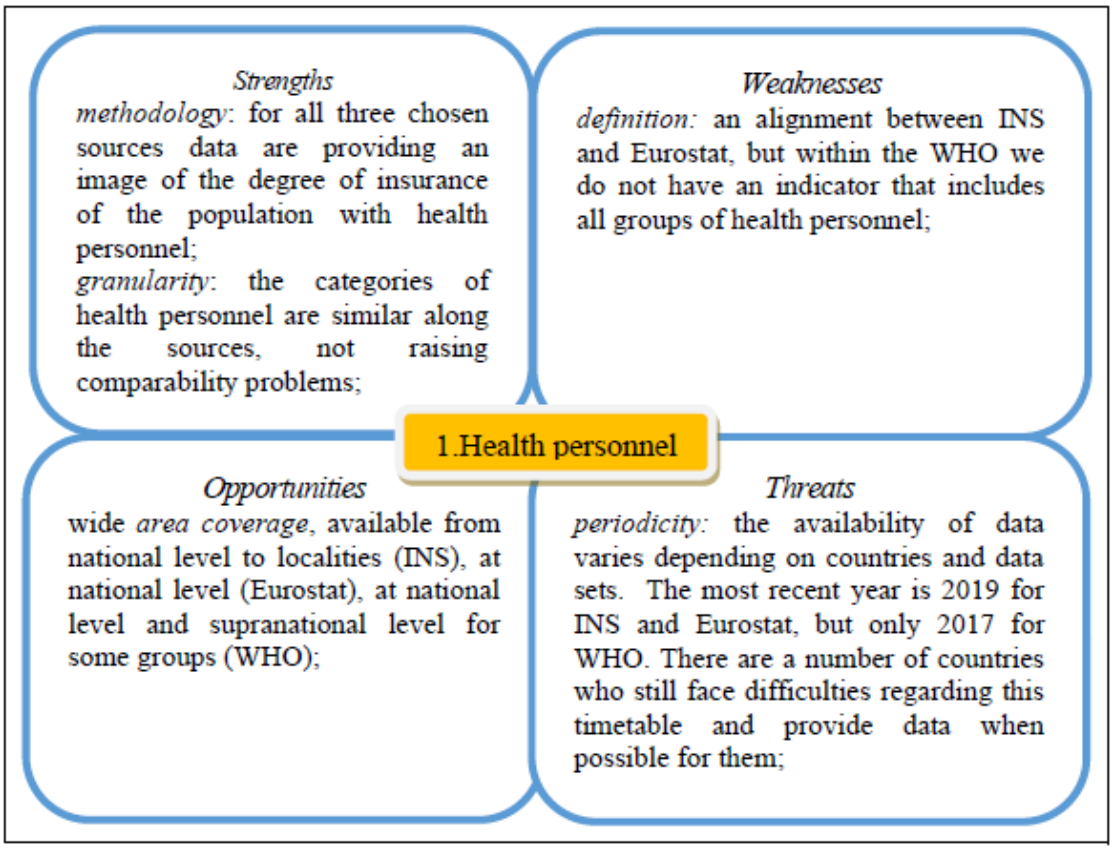

**Figure 1.** SWOT analysis for Health personnel indicator

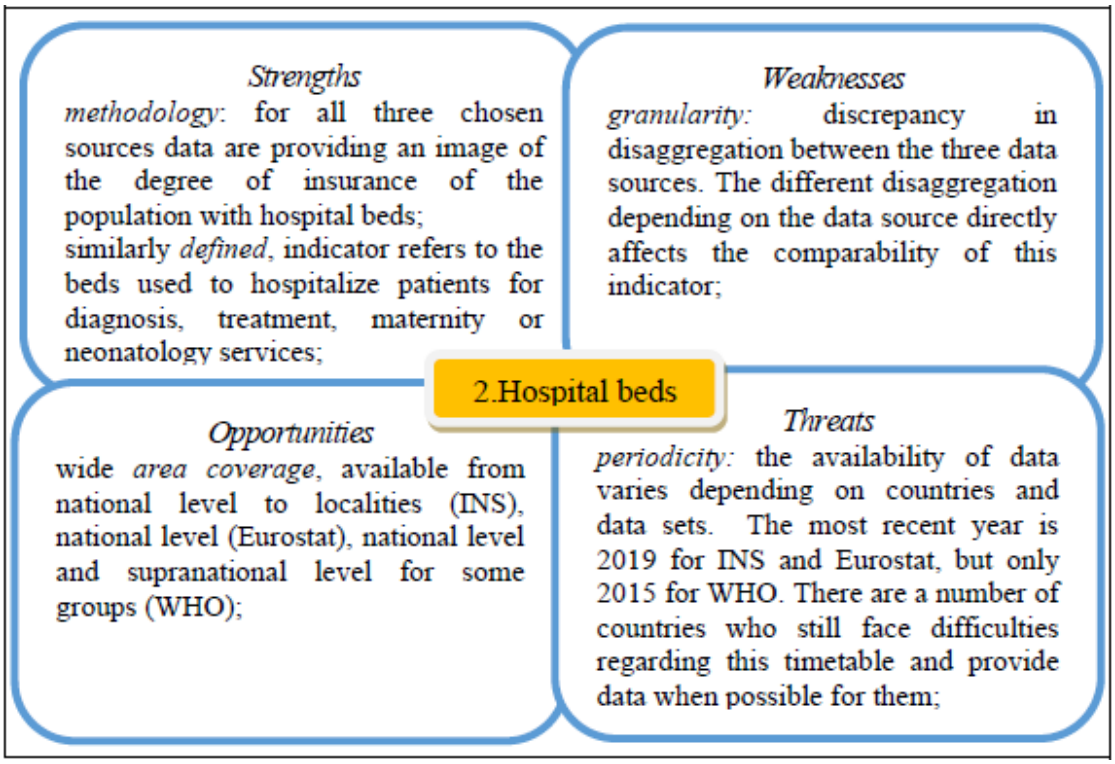


**Figure 2.** SWOT analysis for Hospital beds indicator

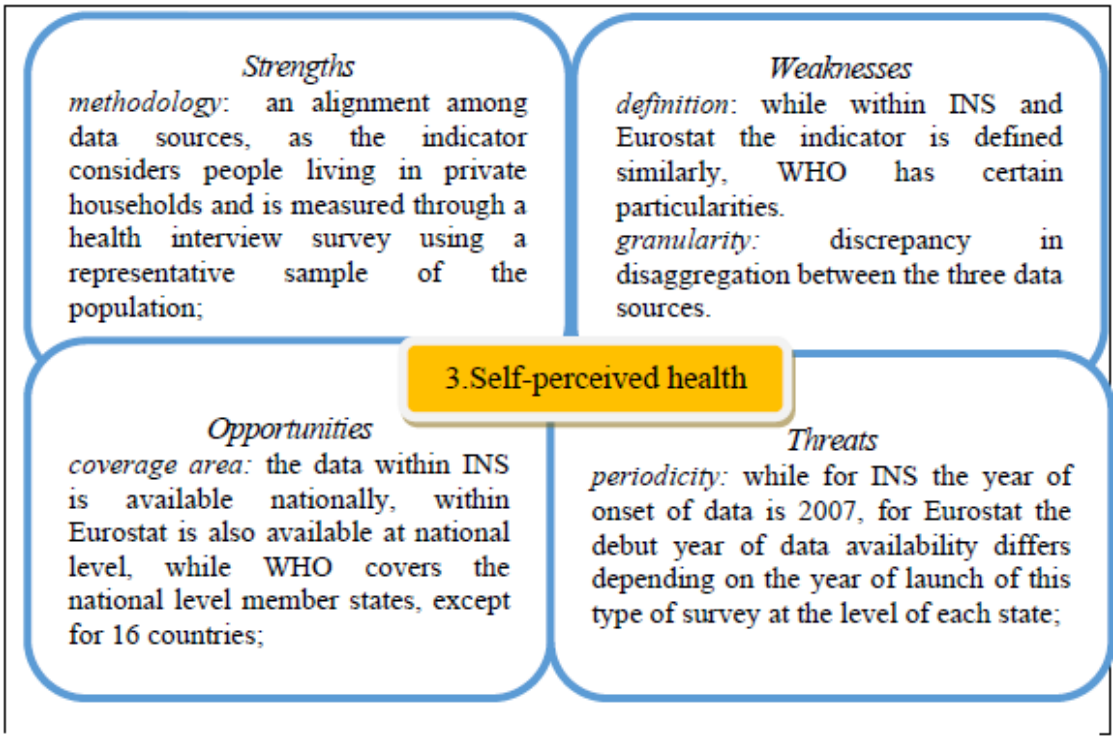

**Figure 3.** SWOT analysis for Self-perceived health indicator

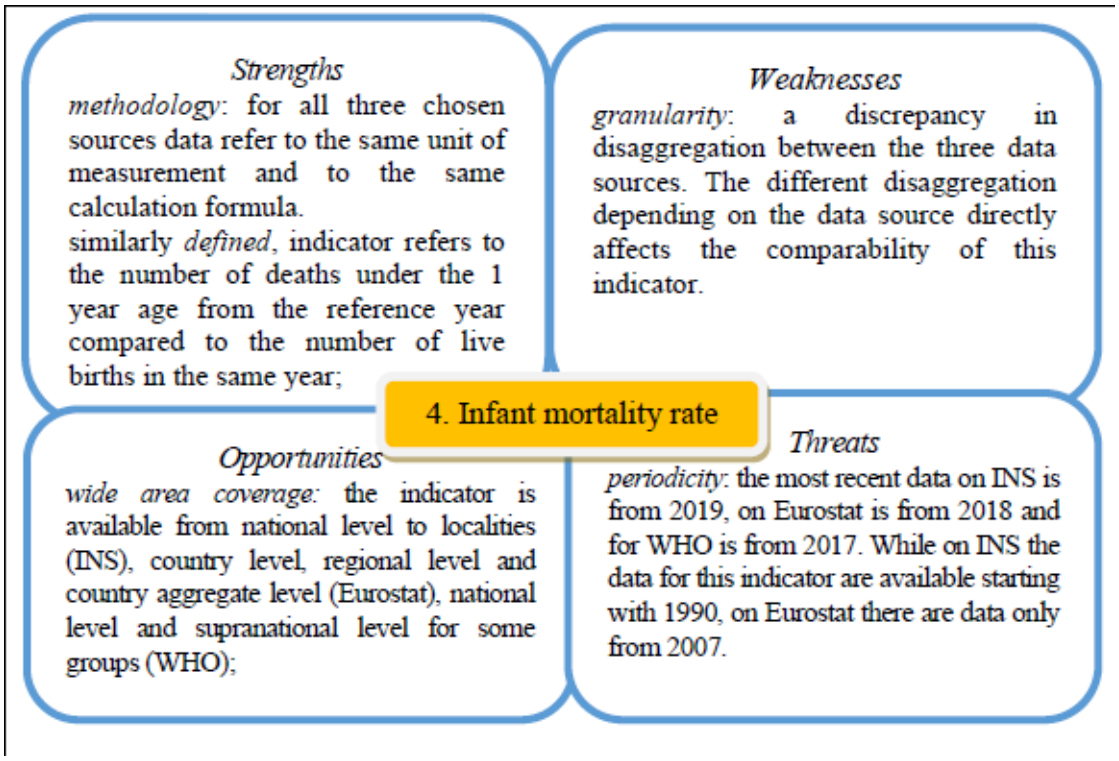


**Figure 4.** SWOT analysis for Infant mortality rate indicator

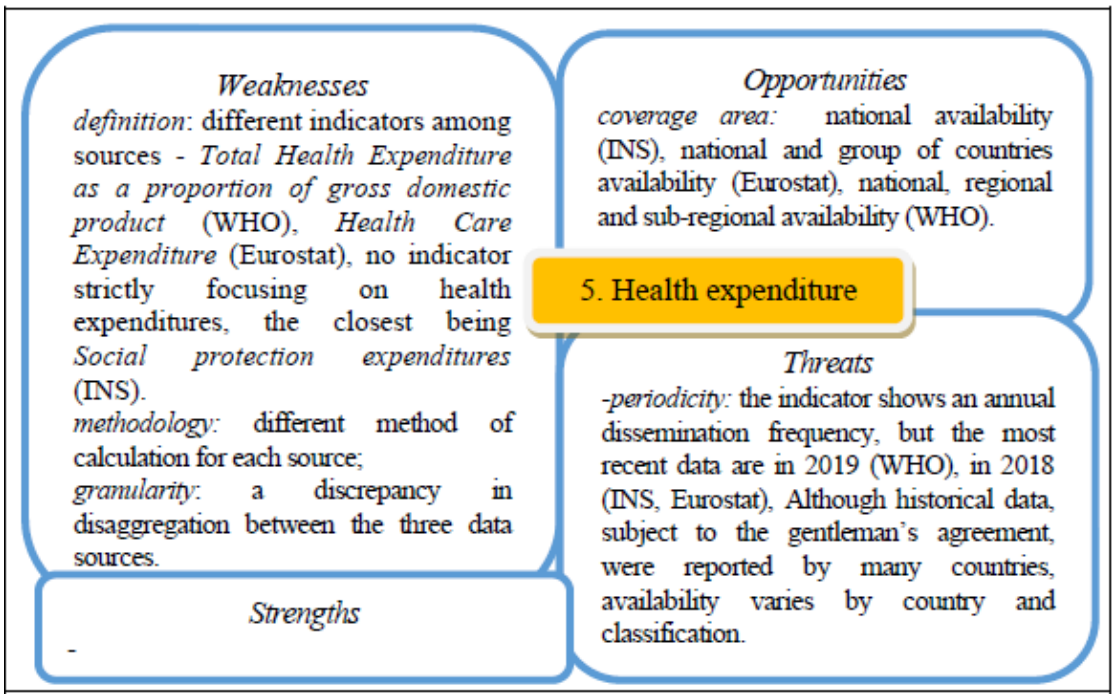

**Figure 5.** SWOT analysis for Health expenditure indicator

## CONCLUSIONS

Despite all the efforts of each institution, there is a limitation of comparability of data between different countries due to the conditioning of the quality of a country's data by the way the provision of healthcare is organized in that country and by the available and collected information by those institutions.

Within the WHO, for OECD Member States, data are taken from the OECD Health database, and for non-OECD countries, the data are reported by the country and may not necessarily meet the common definition of WHO. (World Health Organization, 2020). Within Eurostat, the raw data sent by the national statistical offices is firstly checked to see if the total of one variable is consistent with the breakdown into different variables. Secondly, there are also cross-checks done in order to see if the differences in the same variable are consistent.

The internal coherence of data is ensured by the annual validation processes within Eurostat, while the comparability of data over time is highly verified before dissemination. Consistency is ensured by comparing the statistics with previous year, editing macrodata or by detecting extreme values called outliers. (European Commission, 2020)

There are some countries that are not able to cover all the healthcare providers (as the inclusion of private providers seems particularly difficult) or can only supply data for selective regions. Sometimes the available breakdown does not match the NUTS classification and therefore, regional data cannot be made available. It is reported ongoing work to the Eurostat Working Group on Public Health Statistics in order to increase quality, comparability and coverage. Quality, comparability and coverage are annually discussed together with the OECD and WHO at the trilateral annual meetings on the common questionnaire on non-monetary health statistics, which is the foundation for data collection. (European Commission, 2020)

Combining data sources and computing capacity from different countries is the only way to obtain a sufficient sample size for epidemiological or clinical research, for a better understanding of the natural occurrence of rare diseases and the patient, needed to make progress in diagnosis and treatment of rare diseases.

Only through a convergent effort there can be possible the creation of a common computing and data infrastructure. Pooling scientific and medical expertise is also essential for Europe to benefit from the power of Big Data to provide citizens with early diagnosis of diseases and new personalized treatments.

# REFERENCES


Abualghaib, O., Groce, N., Simeu, N., Carew, M. T. and Mont, D., 2019. Making Visible the Invisible: Why Disability-Disaggregated Data is Vital to "Leave No-One Behind". Sustainability, 11(11), 3091. https://doi.org/10.3390/su11113091

Bartlett, V. L., Dhruva, S. S., Shah, N. D., Ryan, P. and Ross, J. S., 2019. Feasibility of Using Real-World Data to Replicate Clinical Trial Evidence. JAMA Network Open, 2(10), e1912869. https://doi.org/10.1001/jamanetworkopen.2019.12869

Booth, C. M., Karim, S. and Mackillop, W. J., 2019. Real-world data: Towards achieving the achievable in cancer care. Nature Reviews. Clinical Oncology, 16(5), 312-325. https://doi.org/10.1038/s41571-019-0167-7

Ciora, L., & Buligiu, I., 2003. Metode si tehnici de analiza a calitatii datelor. Revista Informatica Economica, nr. 1 (25). http://www.revistaie.ase.ro/content/25/Ciora.pdf

European Commission. (2018). COMMUNICATION FROM THE COMMISSION TO THE EUROPEAN PARLIAMENT, THE COUNCIL, THE EUROPEAN ECONOMIC AND SOCIAL COMMITTEE AND THE COMMITTEE OF THE REGIONS on enabling the digital transformation of health and care in the Digital Single Market; empowering citizens and building a healthier society. [Online] Available on EUR-Lex. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM%3A2018%3A233%3AFIN. [Accessed 15 March 2021]

European Commission's Directorate General of Health and Food Safety. (2017). State of health in the EU. Companion Report 2017 (Health and Food Safety, pp. 1-72). European Commission. [Online] Available at https://ec.europa.eu/health/sites/health/files/state/docs/2017_companion_en.pdf [Accessed 13 March 2021]

European Commission, 2020. Metadata. [Online] Available at: https://ec.europa.eu/eurostat/web/main/data/metadata. [Accessed 10 March 2021]

Lipworth, W., 2019. Real-world Data to Generate Evidence about Healthcare Interventions: The Application of an Ethics Framework for Big Data in Health and Research. Asian Bioethics Review, 11(3), 289–298. https://doi.org/10.1007/s41649-019-00095-1

National Institute of Statistics. (2020). National Institute of Statistics. [Online] Available at Metadata: http://colectaredate.insse.ro/metadata/public.html [Accessed 11 March 2021]

National Institute of Statistics. (2020). National Institute of Statistics. [Online] Available at TEMPO Online: http://statistici.insse.ro:8077/tempo-online/#/pages/tables/insse-table. [Accessed 12 March 2021]

Popa, G., Pavel, r. a., & Căpăţînă, C., 2005. Quality Guidelines for Romanian Official Statistics. National Institute of Statistics.

World Health Organization. (2016). World health statistics 2016: Monitoring health for the sustainable development goals SDGs. World Health Organization. [Online] Available at: https://apps.who.int/iris/handle/10665/206498. [Accessed 5 March 2021]

World Health Organization, 2020. World health statistics 2020: Monitoring health for the SDGs, sustainable development goals. World Health Organization. [Online] Available at: https://apps.who.int/iris/handle/10665/332070. [Accessed 9 February 2021]

World Health Organization, 2020. Global Health Observatory Metadata. [Online] Available at: https://apps.who.int/gho/data/node.metadata. [Accessed 9 February 2021]