INFACT
Joint Action on Health Information

This project is funded by the Health Programme of the European Union

FACT SHEET FOR METHODOLOGICAL GUIDELINES (Ongoing activity)
WP9 "Methodological guidelines to estimate health indicators using linked data and machine learning techniques"

Key outputs

1. Background: The estimation of health indicators from linked data and the application of machine learning techniques is challenging and may require advanced expertise and skills in statistical analysis. There is a need for methodological guidelines, which could systematically guide MSs for using linked data and machine learning techniques to estimate health indicators for public health research.

2. Innovative tools to improve the current EU-HIS: These guidelines were developed for the first time to adopt, develop and compare new methods/techniques using linked data and machine learning techniques. In turn, this would help to improve the methodological approaches for public health research.

3. Most relevant results: These guidelines contain the following seven important contents: 1. Rational and objective of the study (i.e., research question), 2. Study design, 3. Study population/sample, 4. Linked data sources, 5. Study outcomes, 6. Data preparation and 7. Data analysis. These aspects would be described by using the examples of methodological studies.

4. Feasibility of being integrated in HIS and translated into policies: These guidelines would support to improve the quality and comparability of health information. The estimated health indicators would guide policy process.

Background and Rationale

The capacity to use data linkage and/or the use of artificial intelligence to estimate and predict health indicators varies across EU-MSs (European Member States)[1]. However, the estimation of health indicators from linked administrative data is challenging due to several reasons such as variability in data sources and data collection methods, availability of a large number of variables, lack of skills and capacity to link and analyze big data[2]. To our knowledge, there are no methodological guidelines available, which could systematically guide MSs for using linked data and machine learning techniques to estimate health indicators. Therefore, the InfAct project has proposed to develop these guidelines, which could guide those MSs who are planning to estimate health indicators using linked data and artificial intelligence with new methods/techniques. The main objective of this study is to develop the methodological guidelines for studies to guide European countries using linked data and artificial intelligence with new methods/techniques.

Proposal

We have performed four following steps systematically to develop the methodological guidelines: 1. scientific literature review; 2. development of generic method; 3.

identification of inspiring examples or best practices from European countries; and 4. validation by a panel of experts.

Expected Results: These guidelines contain the following seven important contents: 1. Rational and objective of the study (i.e., research question), 2. Rational for the selection of a study design, 3. Selection of study population/sample, 4. Linked data sources available, 5. Defining the study outcomes, 6. Data preparation and 7. Data analysis. We have described these aspects with examples of different methodological studies.

Recommendations for sustainability

These guidelines aim to adopt, develop and compare new methods/techniques using linked data and machine learning techniques for public health research studies.

We proposed the following recommendations based on these guidelines:

- Rational selection of the study design using linked data is important and may avoid certain methodological limitations.
- Standard methods for data collection should be implemented in a HIS.
- Routinely data collected from various administrative sources should improve their quality concerning to the completeness of information.
- Data related to employment, education, occupation and socioeconomic status should be readily available/accessible for analysis related to the health status.
- Specific mandates to ensure data availability/access/capture and safe storage should be an integral part of a national/regional health information system[1]
- More collaborations among Member States for an exchange of inspiring examples/best practices in using linked data and machine-learning approaches are needed in the future among European countries.
- Joint country studies on using machine-learning techniques for public health research are needed.
- Better approaches to translate estimated health indicators for health policy are required.

Such guidelines need to be revised after inputs from experts.

References

1. Haneef R, Delnord M, Vernay M, et al. Innovative use of data sources: a cross-sectional study of data linkage and artificial intelligence practices across European countries. *Archives of Public Health.* 2020;78(1):55.
2. Bradley CJ, Penberthy L, Devers KJ, Holden DJ. Health Services Research and Data Linkages: Issues, Methods, and Directions for the Future. *Health Services Research.* 2010;45(5p2):1468-1488.