Using Large Cohort Data to Strengthen Information-Based Public Health Policies: An Appeal to Portuguese Authorities

Utilização de Grandes Bases de Dados Longitudinais para Fortalecer as Políticas de Saúde Pública Baseadas em Informação: Um Apelo às Autoridades Portuguesas

Dear Editor,

Everyday millions of bytes of data are generated by routine activities of healthcare systems. Nevertheless, health researchers opt for primary data collection. This decision is often justified by the low quality of existing data, legal and bureaucratic processes impeding timely access, and/or important discrepancies between the needed data and the data that is available.

In Portugal, national data is siloed. Health data is collected in different datasets, is not interoperable or linked, with one identifier per dataset and several data 'owners'. Therefore, health data is not being used to anywhere near its full potential, leading to 'data wastage', hampered research and development (R&D) competitiveness, increased costs, and potentially slower responses to emerging health problems.

We argue that improved, ethical usage of national data registries to support public health decisions is paramount to Portuguese R&D strategy, competitiveness, and building an epidemiological milieu. Some of the research priorities funded through national and international research agencies can be addressed through the analysis of this data, allowing for more complex analyses, faster results, and a lower cost.

The existing unclear, unstructured processes for accessing data are a great disadvantage for Portuguese researchers. Even in cases where researchers are granted access, not all the necessary data to test the hypothesis are provided (e.g., date of birth, borough/parish, sex). Some can be denied due to concerns regarding confidentiality/anonymity of subjects in the datasets, making testing for potential confounders in the data *de facto* impossible)

Having clear paths to access big datasets would increase the relevance, efficiency and usefulness of the data collection, and its quality. Having several dataset users can help in identifying errors and pitfalls, bridging information needs and demand for data.

The Nordic countries are an example to follow, where national registries and the use of a single unique identifier enable the development of large cohort studies involving millions of participants,[1-3] with adequate censoring for death and migration, based on individual-level data collected for administrative purposes, either individually or per groups of countries.[4]

In Denmark, for instance, researchers can access anonymized raw data, use the Statistics Denmark's online research machine to analyze large datasets and transfer the results of the analysis afterwards.[5]

It is our hope that the Portuguese authorities, including the Ministries of Health, Education, Science and Technology will be inspired by these examples and derive a strategy to increase access and usefulness of the national registries, otherwise, Portuguese researchers and the country will lose its R&D competitiveness.

AUTHOR CONTRIBUTIONS

IF: Study design and writing of the manuscript.

FSB, CSB, PF, HB: Study design and critical review of the manuscript.

All authors approved the final version to be published.

PROTECTION OF HUMANS AND ANIMALS

The authors declare that the procedures were followed according to the regulations established by the Clinical Research and Ethics Committee and to the Helsinki Declaration of the World Medical Association updated in 2013.

DATA CONFIDENTIALITY

The authors declare having followed the protocols in use at their working center regarding patients' data publication.

COMPETING INTERESTS

The authors have declared that no competing interests exist.

FUNDING SOURCES

REFERENCES

1. Rod NH, Andersen TO, Severinsen ER, Sejling C, Dissing A, Pham VT, et al. Cohort profile: The SmartSleep Study, Denmark, combining evidence from survey, clinical and tracking data. BMJ Open. 2023;13:e063588.
2. Karlstad Ø, Hovi P, Husby A, Härkänen T, Selmer RM, Pihlström N, et al. SARS-CoV-2 vaccination and myocarditis in a Nordic cohort study of 23 million residents. JAMA Cardiol. 2022;7:600-12.
3. Maret-Ouda J, Tao W, Wahlin K, Lagergren J. Nordic registry-based cohort studies: possibilities and pitfalls when combining Nordic registry data. Scand J Public Health. 2017;45:s14-9.
4. Laugesen K, Ludvigsson JF, Schmidt M, Gissler M, Valdimarsdottir UA, Lunde A, et al. Nordic health registry-based research: a review of health care systems and key registries. Clin Epidemiol. 2021;13:533-54.
5. Statistics Denmark. Data for research. [cited 2024 Jan 23]. Available from: https://www.dst.dk/en/TilSalg/Forskningsservice.

Inês FRONTEIRA✉*[1,2], Frederik SCHALTZ-BUCHHOLZER[3], Christine STABELL BENN[3,4], Paulo FERRINHO*[2,5], Henrique BARROS*[6,7,8]

* Equally shared authorship position
1. Comprehensive Health Research Centre (CHRC). NOVA Medical School. Universidade NOVA de Lisboa. Lisbon. Portugal.
2. Associação Portuguesa de Epidemiologia. Porto. Portugal.
3. Odense Patient data Explorative Network (OPEN). Bandim Health Project. Institute of Clinical Research. Odense University Hospital/University of Southern Denmark. Odense. Denmark.
4. Danish Institute for Advanced Study. University of Southern Denmark. Odense. Denmark.
5. Global Health and Tropical Medicine. Instituto de Higiene e Medicina. Universidade NOVA de Lisboa. Lisbon. Portugal.
6. Epidemiology Research Unit (EPIUnit). Instituto de Saúde Pública. Universidade do Porto. Porto. Portugal.
7. Laboratório para a Investigação Integrativa e Translacional em Saúde Populacional (ITR). Porto. Portugal.
8. Departamento de Ciências da Saúde Pública e Forenses e Educação Médica. Faculdade de Medicina. Universidade do Porto. Porto. Portugal.
✉ **Autor correspondente:** Inês Fronteira. ines.fronteira@ensp.unl.pt

https://doi.org/10.20344/amp.21329