The European Journal of Psychiatry 38 (2024) 100242

THE EUROPEAN JOURNAL OF PSYCHIATRY

www.elsevier.es/ejpsy

EDITORIAL

Do we have data Diogenes in research? Seven questions and seven suggestions to identify and manage it for the sake of participants and the advancement of our research field

Rosa Ayesa-Ariola[a,b,*], **Marta Rapado-Castro**[c,d]

[a] *Department of Psychiatry, Valdecilla Biomedical Research Institute (HUMV-IDIVAL), Santander, Spain*
[b] *Biomedical Research Networking Center for Mental Health (CIBERSAM), Madrid, Spain*
[c] *Department of Child and Adolescent Psychiatry, Institute of Psychiatry and Mental Health, Hospital General Universitario Gregorio Marañón, School of Medicine, Universidad Complutense, IiSGM, ISCIII, Madrid, Spain*
[d] *Melbourne Neuropsychiatry Centre, Department of Psychiatry, The University of Melbourne and Melbourne Health, 161 Barry Street, Carlton South, Victoria, Australia*

Received 4 October 2023; accepted 14 October 2023

Congresses and scientific meetings have those moments when ideas arise that may not occur in other circumstances. They can be given by listening to a conference, commenting over coffee, or in the leisure moments in which the talk about life and scientific aspects continues. It was in one of those moments of the past Congress of the Schizophrenia International Research Society (SIRS 2023) in Toronto when we had a moment of epiphany and this letter was born.

We were in a thrift store, packed with antiques, old toys, books, action figures, clothes and all sort of things that spreaded over the shelves, floors, and everywhere around. It was then when Dr. Rapado-Castro formulated the phrase: "I think we have data Diogenes in research". And just like that, as we were discussing about how that notion was related so much to our experience as senior researchers, and how we felt about that we thought the existence of data Diogenes in research was such a real and interesting appreciation that we had to share our concern with the scientific community.

Raise your hand if your research group has always used or exploited all the data collected during the course or the development of a research project. The main research groups that we have worked with, visited or just known throughout our clinical research careers in mental health did not. Although for many years we have been collecting data and had proposed hypotheses and being able to contrast them, we are still having a large amount of data that needs to be tested. Even so, we are still applying for funding and writing projects in which it is proposed to collect a large amount of data again. Likely some of this data will never be used/analyzed due to restrictions in funding, specialized personnel, time dedicated to teaching, supervision or to research in a clinical setting or discontinuation of the former: blood and other biological samples that can remain preserved in biobanks for years with the consequent storage costs and the inevitable degradation; variables extracted from the administration of sociodemographic and clinical scales collected during long interviews; results of neuropsychological batteries that involve many hours of administration and correction work; neuroimaging data that is very expensive and complex to obtain and requires specific competences and expertise to be analyzed. Collecting all this information requires great amount of time and generosity

* Corresponding author at: Department of Psychiatry, Valdecilla Biomedical Research Institute (HUMV-IDIVAL), Santander, Spain.
E-mail address: rayesa@idival.org (R. Ayesa-Ariola).

https://doi.org/10.1016/j.ejpsy.2023.100242

on the part of the study participants, who also lend themselves to being re-contacted longitudinally to repeat the entire evaluation process in follow-up visits at 1, 3 or 6 months and then at 1, 2, 3, 5, 10, 15 and even 20 and 30 years. Most of their data will then remain stored in complex databases, sometimes forgotten in the natural change of personnel/researchers involved in the projects. In this context, the term "data Diogenes" seems very appropriate to describe our behaviors of continuing to accumulate something that is not used and can become useless over time, compulsively repeating these behaviors over and over again for the sake of having all areas or research covered at expenses to the time and effort of all the parties involved.

Our comment seeks to be a call to reflection for those colleagues who share this concern and think change is possible in the preparation, evaluation and approval of research projects. Perhaps asking ourselves a few questions might help in this process.

1. Is there a realistic timeframe for using the data while they are still relevant and valid? —This timeframe depends on the nature of the data and the specific degradation factors (i.e. genetics or neuroimaging data influenced by technological advances and preservation practices). In addition, if the data represents a specific time period or is tied to a particular context, its usefulness may diminish over time as circumstances change. This might be the case for data gathered during the COVID-pandemic, which might require being shortly analyzed and disseminated within a very concrete timeframe.
2. Who is going to analyze the data and publish the results (both positive and negative findings)? -There are databases or samples that have never been used/analyzed; there are also unpublished negative results that might be of interest. PhD, medical, or mental health students, registrars, other neuroscientist in our team, as well as other external collaborators might be interested in pursuing our research aims or able to contribute with their own time or expertise.
3. Have I already used/analyzed a large part of the baseline data and is it time to request a longitudinal project? Can I apply for funding to reuse some of the data I already have and analyze it using a different approach? Can I formulate different hypothesis using the data I already have? -Longitudinal projects are justifiable when the research questions necessitate the examination of processes, changes over time, long-term effects, or causal relationships. Other research questions can be easily answered using cross-sectional data, as this might inform on specific issues relevant for the present clinical practice or inform about the subsequent need for longitudinal follow up in a specific group of participants.
4. Can I use scales/batteries/protocols to optimize the assessments by shortening the evaluation times and the number of variables gathered from the participants in the studies? -Think about limiting the longitudinal collection of data to those variables that have been significant/of interest in the baseline evaluations.
5. Can I share the data with other researchers/groups that can pool it with their samples or reanalyze it using a different relevant approach? —Research Consortia on a particular topic are becoming important resources to yield relevant results by increasing the statistical power/sample sizes while aiding to form data cleaning/reusing data teams.
6. Is it an ethical behavior to continue collecting and accumulating data without using them? - Out of respect for the patients and participants in the studies in general, it might be interesting to have a sit and try to answer the questions with the information we already have.
7. Can I propose new projects/hypotheses using the data that is already available? -It might be possible to propose new projects or hypotheses using existing data, following the principles of "recover, recycle, and reuse".

In many cases, existing datasets can be a valuable resource for conducting new analyses or exploring additional research questions. Here's how we can approach the proposals for new projects or test of hypotheses using already available data:

1. Data evaluation: Start by evaluating the existing data to determine its relevance and suitability for your proposed research project. Consider factors such as the quality of the data, the type of variables that have been collected, the sample size, and the time frame that has been covered. Ensure that the existing data aligns with the specific research questions or hypotheses you want to test or to explore.
2. Identify gaps and opportunities: Examine the existing data to identify any gaps or areas of research that have not yet been fully explored. Look for patterns, relationships, or variables that could be further analyzed, combined or be approached on a new fashion to address new research questions. Consider potential subgroups or subsets within the data that may offer unique insights.
3. Reframe research questions or hypotheses: Based on the evaluation of the existing data and the identified gaps, reframe or develop new research questions or hypotheses that can be explored using the available data. Consider how your newly proposed project aligns with the themes of "recover, recycle, reuse" and how it may contribute to knowledge generation or addressing important issues maximizing current resources.
4. Design appropriate analyses: Determine the appropriate statistical or analytical techniques to be applied to the existing data, to answer your research questions or to test your hypotheses. This may involve conducting descriptive analyses, inferential statistics, regression models, or other relevant methods depending on the nature of the data and on your research objectives.
5. Validate and interpret your findings: Once you have conducted the analyses, validate your findings and interpret the results within the context of the existing data. Discuss the implications of your findings and how they contribute to the understanding of the research area or address the specific issues.
6. Consider limitations and future directions: Acknowledge any limitations of the existing data and of your proposed new project. Discuss potential avenues for future research, including the need for additional data collection or the integration of other data sources to enhance the comprehensiveness of the analysis.

7. Remember to properly cite and acknowledge the source of the existing data in your new project or hypothesis proposal. Ethical considerations should be followed when reusing data, including obtaining necessary permissions or adhering to any restrictions or guidelines set by the original data collectors or providers.

We hope these suggestions may assist to optimize resources in research and counteract our data Diogenes in research. We firmly believe that it is possible to contribute to the body of knowledge and shed light on important topics by leveraging existing data through innovative approaches and new research questions.

Conflicts of Interest

The authors declare no conflict of interest.

Ethical considerations

Maintenance of the highest level of objectivity throughout the manuscript.

Funding

This research received no external funding. Dr. Ayesa-Arriola is funded by a Miguel Servet contract from the Carlos III Health Institute (isc), carried out on Fundación Instituto de Investigación Marqués de Valdecilla (IDIVAL). Her research was partially supported by the Spanish Ministry of Science and Innovation, Instituto de Salud Carlos III, ISCIII, (PI17/00221, PI20/00066, CNS2022-136110). Dr. Rapado-Castro is a Ramon y Cajal Research Fellow (RYC-2017-23144), Spanish Ministry of Science, Innovation and Universities and was supported by a NARSAD independent investigator grant (No. 24628) from the Brain & Behavior Research Foundation. MR-C was partially supported by the Spanish Ministry of Science and Innovation, Instituto de Salud Carlos III, ISCIII, (PI18/00753, PI21/00701), CIBER -Consorcio Centro de Investigación Biomédica en Red- (CB/07/09/0023), co-financed by the European Union and ERDF Funds from the European Commission, “A way of making Europe”, financed by the European Union - NextGenerationEU (PMP21/00051), Madrid Regional Government (B2017/BMD-3740 AGES-CM-2), European Union Structural Funds, EU Seventh Framework Program, H2020 Program, and Horizon Europe, Fundación Familia Alonso, and Fundación Alicia Koplowitz.