Evaluating the Replicability of MIDAS Studies: A Pilot Study

**Introduction**

Replicability of research studies is a fundamental part of the scientific process.

Recently, various studies have highlighted the lack of replicability and reproducibility of research in certain scientific domains, such as computer science [refs]. Replicability, repeatability, and reproducibility are related concepts, but have each their own specific definition. Replicability is..... <etc>. The concept of Open Science has emerged to address limited replicability of scientific studies, through guidelines and infrastructure for sharing of data, software, source code, and workflows. Previously, replicability studies have been conducted in computer science <refs> psychology [4], ecology and evolution [5–7]; however, to our knowledge, replicability studies have not been conducted for infectious disease modeling research. The FAIR principles have emerged as a guide towards re-usable digital objects.

<remove>

t. Even when the raw data and code used to produce the study results are present, the quality of data documentation often impedes future replication and reproduction attempts. To optimize the accessibility of data, the Findable, Accessible, Interoperable, and Reusable principles were created to guide researchers in data management and publishing [3]. In this study, we will use the FAIR principles as a guide to evaluate the replicability of publications written by Models of Infectious Disease Agent Study (MIDAS) members.

<remove>

The purpose of this pilot study is to assess whether evaluating the replicability of studies conducted by MIDAS members is feasible and to determine the design of a full-scale replicability study. The specific feasibility objectives of this study are to determine if: 1) a study can be categorized as replicable given the published information, and 2) Occam software can be utilized to replicate study results when raw code and data are available. We will determine that it is feasible to carry on the full-scale study if the surveyed MIDAS publications meet these objectives.

**Study Objectives**

1. Evaluate the replicability of studies published by the MIDAS members.
2. When raw data and code are available, use Occam to replicate the study results.

**Methods**

Five studies were selected from a compiled list of first and last authored publications for each of 175 MIDAS members, as of <date> . For each study, raw data availability, whether the data were available in a single repository or collected from multiple sources, format of the data, and software used for data transformation were documented. Additionally, presence of code and the software used to generate the results were also recorded. When raw data and code were available, an attempt was made to upload the information into Occam to replicate the study results.

**Results**

Raw data was publicly available for three of the five studies [8–10]. Of the three studies with available raw data, two compiled the data within a single repository with multiple formats [8–9]; the third study referenced multiple sources where data in various formats were pulled [10]. Code used to run the full analysis of the study was available for two of the publications [8–9]. One study provided code used to produce the results but did not provide information on data transformation methods [10]. The fourth study stated the software used for data transformation and analysis and provided a link for the raw data and code; however, neither was available when the link was followed [11]. The last study did not provide any raw data or code that was used in their analysis [12] (Table 1).

**Table 1. Raw data and code availability assessment.**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Study Number | Data Availability | Data Available in Single Repository | Data Format | Software for Data Transformation | Code Availability | Analysis Software | Study Reference |
| 1 | Yes | Yes | csv or rda | R | Yes | R | [8] |
| 2 | Yes | Yes | JSON | R, coffee, java script, or python | Yes | R | [9] |
| 3 | Yes | No | csv, pdf, excel, and docx |  | Yes | C++ | [10] |
| 4 | No |  |  | Perl, MATLAB, and R | No | C++ | [11] |
| 5 | No |  |  |  | No |  | [12] |

**Discussion**

We were able to evaluate whether each of the five studies were replicable based on the availability of the raw data and code. Based on the assessment in Table 1, only two had published the necessary components to be categorized as replicable [7–8]. While the fourth study provided references for where the raw data were collected and the software used to analyze the results, they did not provide the code used for data transformation and parameter estimations [11]. For our main study to be feasible, it would be appropriate to create a replicability scale that incorporates the FAIR principles to better categorize each study. Additionally, in the pilot study we were unable to use Occam because it is currently unable to analyze data with R. In order for the main study to be feasible, Occam needs to be expanded to accept code from a larger variety of statistical software programs.

**References**

1. Tenopir C, Allard S, Douglass K, Aydinoglu AU, Wu L, Read E, et al. Data sharing by scientists: practices and perceptions. PloS one. 2011;6(6):e21101. Epub 2011/07/09. doi: 10.1371/journal.pone.0021101. PubMed PMID: 21738610; PubMed Central PMCID: PMCPMC3126798.

2. Fecher B, Friesike S, Hebing M. What drives academic data sharing? PloS one. 2015;10(2):e0118053. Epub 2015/02/26. doi: 10.1371/journal.pone.0118053. PubMed PMID: 25714752; PubMed Central PMCID: PMCPMC4340811.

3. Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, et al. The FAIR Guiding Principles for scientific data management and stewardship. Scientific data. 2016;3:160018. Epub 2016/03/16. doi: 10.1038/sdata.2016.18. PubMed PMID: 26978244; PubMed Central PMCID: PMCPMC4792175.

4. PSYCHOLOGY. Estimating the reproducibility of psychological science. Science (New York, NY). 2015;349(6251):aac4716. Epub 2015/09/01. doi: 10.1126/science.aac4716. PubMed PMID: 26315443.

5. Jennions MD, Moller AP. Publication bias in ecology and evolution: an empirical assessment using the 'trim and fill' method. Biological reviews of the Cambridge Philosophical Society. 2002;77(2):211-22. Epub 2002/06/12. PubMed PMID: 12056747.

6. Fidler F, Chee YE, Wintle BC, Burgman MA, McCarthy MA, Gordon A. Metaresearch for Evaluating Reproducibility in Ecology and Evolution. Bioscience. 2017;67(3):282-9. Epub 2017/06/10. doi: 10.1093/biosci/biw159. PubMed PMID: 28596617; PubMed Central PMCID: PMCPMC5384162.

7. Roche DG, Kruuk LE, Lanfear R, Binning SA. Public Data Archiving in Ecology and Evolution: How Well Are We Doing? PLoS biology. 2015;13(11):e1002295. Epub 2015/11/12. doi: 10.1371/journal.pbio.1002295. PubMed PMID: 26556502; PubMed Central PMCID: PMCPMC4640582.

8. Reich NG, Lauer SA, Sakrejda K, Iamsirithaworn S, Hinjoy S, Suangtho P, et al. Challenges in Real-Time Prediction of Infectious Disease: A Case Study of Dengue in Thailand. PLoS neglected tropical diseases. 2016;10(6):e0004761. Epub 2016/06/16. doi: 10.1371/journal.pntd.0004761. PubMed PMID: 27304062; PubMed Central PMCID: PMCPMC4909288.

9. Brooks LC, Farrow DC, Hyun S, Tibshirani RJ, Rosenfeld R. Nonmechanistic forecasts of seasonal influenza with iterative one-week-ahead distributions. PLoS computational biology. 2018;14(6):e1006134. Epub 2018/06/16. doi: 10.1371/journal.pcbi.1006134. PubMed PMID: 29906286; PubMed Central PMCID: PMCPMC6034894.

10. Arnold BJ, Gutmann MU, Grad YH, Sheppard SK, Corander J, Lipsitch M, et al. Weak Epistasis May Drive Adaptation in Recombining Bacteria. Genetics. 2018;208(3):1247-60. Epub 2018/01/14. doi: 10.1534/genetics.117.300662. PubMed PMID: 29330348; PubMed Central PMCID: PMCPMC5844334.

11. Espana G, Grefenstette J, Perkins A, Torres C, Campo Carey A, Diaz H, et al. Exploring scenarios of chikungunya mitigation with a data-driven agent-based model of the 2014-2016 outbreak in Colombia. Scientific reports. 2018;8(1):12201. Epub 2018/08/17. doi: 10.1038/s41598-018-30647-8. PubMed PMID: 30111778; PubMed Central PMCID: PMCPMC6093909.

12. Barton HD, Rohani P, Stallknecht DE, Brown J, Drake JM. Subtype diversity and reassortment potential for co-circulating avian influenza viruses at a diversity hot spot. The Journal of animal ecology. 2014;83(3):566-75. Epub 2013/10/30. doi: 10.1111/1365-2656.12167. PubMed PMID: 24164627; PubMed Central PMCID: PMCPMC4000580.