Transparency in infectious disease research

Emmanuel A. Zavalis1, 2†, Despina Contopoulos-Ioannidis3†, John P.A. Ioannidis2,4\*†

1Department of Learning Informatics Management and Ethics, Karolinska Institutet, Stockholm, Sweden

2Meta-Research Innovation Center at Stanford (METRICS), Stanford University, Stanford, CA, USA

3Division of Infectious Diseases, Department of Pediatrics, Stanford University School of Medicine, Stanford, CA, USA

4Stanford Prevention Research Center, Department of Medicine, and Department of Epidemiology and Population Health, Stanford University School of Medicine, Stanford, CA, USA

**\* Correspondence:**Corresponding Author  
[email@uni.edu](mailto:email@uni.edu)

†These authors contributed equally to this work and share first authorship

Keywords: Infectious Disease, Transparency, Reproducibility, Rigor, Meta-Research, Meta-epidemiology

# Material and Methods

The study will be a meta-epidemiological assessment of transparency indicators within the most cited infectious disease journals.

## Study sample

This study will be a meta-epidemiological survey of transparency indicators present the literature within the 9 journals with the most citations in infectious disease research from the InCites Journal Citation Reports (JCR)(1) indexed in the PubMed Central Open Access (PMC OA) subset of PubMed (The journals being: *BMC Infectious Diseases*, *International Journal of Infectious Diseases*, *Journal of Infectious Diseases*, *Journal of Antimicrobial Chemotherapy*, *Emerging Infectious Diseases*, *Clinical Microbiology and Infection*, *Infection and Immunity*, *Clinical Infectious Diseases* and *Lancet Infectious Diseases* in no particular order)

### Search query

The search was performed on PubMed on the 14th of September 2022 using the query: (pubmed pmc open access[filter]) AND (1058-4838[journal] OR 0022-1899[journal] OR 1473-3099[journal] OR 0019-9567[Journal] OR 1080-6040[Journal] OR 0305-7453[journal] OR 1198-743X[journal] OR 1471-2334[journal] OR 1201-9712[journal]) AND (journal article[pt]) NOT (editorial[pt] OR letter[pt] OR clinical trial protocol[pt]) AND (fha[Filter]) AND (english[Filter]).

## Data extraction

### Text mining

For each eligible article we will use PubMed to extract information on metadata that includes PMID, PMCID, publication year, journal name and the R package rtransparent(2) to extract the following transparency indicators: (i) code sharing (ii) data sharing (iii) (pre-)registration, (iv) COI and (v) funding statements. rtransparent searches through the full text of the papers for specific words or phrases that strongly suggest that the aforementioned transparency indicators are present in that particular paper. The program uses regular expressions to adjust for variations in expressions.

### Manually extracted parameters

From a random sample of 10% of the publications of each year we intend to validate the algorithm for each transparency indicator by:

* Validating publications classified as not containing the transparency indicator: i.e. if a paper is classified as non-transparent we will assess the validity of this classification.
* Validating publications classified as containing the transparency indicator: i.e. if a paper is classified as transparent for the same parameter we will examine the validity of this classification.
* We will also assess the proportion of Conflict of Interest and Funding disclosures that report that there is no information to report.

For the random sample of publications, we will also extract:

* The studied pathogen
* The study design (classified as a non-clinical trial experimental, clinical trial, observational, or a review)
* Anything else we would like to extract?

### Adjusted estimates of prevalence of the indicators

The corrected proportion of publications satisfying an indicator will beobtained by , where is the uncorrected proportion detected by the automated algorithm, TP will be the proportion of true positives (proportion of those manually verified to satisfy the indicator among those identified by the algorithm as satisfying the indicator, and FN will be the proportion of false negatives (proportion of those manually found to satisfy the indicator among those categorized by the algorithm not to satisfy the indicator).

## Comparisons and statistical analysis

We will consider three primary comparisons that will be conducted using Fisher’s exact tests.

* All publications in 2019 to all in 2021 (to assess if there is improvement over time)
* All non-COVID-19 publications in 2019 to the non-COVID-19 publications in 2021 (to assess if there was improvement over time for non-COVID-19 publications)
* 2021 COVID-19 publications to 2021 non-COVID-19 ones (to assess if COVID-19 papers differ in transparency indicators versus non-COVID-19 papers).

The secondary comparisons will be of:

* The transparency of the research of specific pathogens (comparing the five most common pathogens of our sample and the rest combined to each other)
* Journals’ transparency
* Study design transparency

I don’t think a regression would add much to this as it makes no sense to examine as this study is primarily descriptive, what do you think?

For the statistical analysis we will use a significance threshold of 0.005.

The code needed for the analysis of our data will make use of R 4.2.1 and Python 3.9.7.

# Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

# Funding

Details of all funding sources should be provided, including grant numbers if applicable. Please ensure to add all necessary funding information, as after publication this is no longer possible.

# Acknowledgments

This is a short text to acknowledge the contributions of specific colleagues, institutions, or agencies that aided the efforts of the authors.

# Data Availability Statement

The code and data generated and analyzed for this study will be found in the [OSF REPOSITORY] [LINK] (for the datasets) [GITHUB REPOSITORY][LINK](for the code used).