Journal of Clinical Epidemiology 83 (2017) 57–64

Journal of Clinical Epidemiology

Indirect evidence of reporting biases was found in a survey of medical research studies

Loai N. Albarqouni[a,*], José A. López-López[b], Julian P.T. Higgins[b]

[a]*Centre for Research in Evidence-Based Practice (CREBP), Faculty of Health Science and Medicine, Bond University, 14 University Dr, Robina, Gold Coast, Queensland 4226, Australia*

[b]*School of Social and Community Medicine, University of Bristol, Canynge Hall, 39 Whatley Road, Bristol BS8 2PS, UK*

Accepted 28 November 2016; Published online 11 January 2017

Abstract

Objectives: To explore indirect evidence of reporting biases by examining the distribution of *P*-values reported in published medical articles and to compare *P*-values distributions across different contexts.

Study Design and Setting: We selected a random sample ($N = 1{,}500$) of articles published in PubMed in March 2014. We extracted information on study type, design, medical discipline, and *P*-values for the first reported outcome and primary outcome (if specified) from each article. We plotted the *P*-values transformed to the z-score scale and used caliper tests to investigate threshold effects.

Results: Out of the 1,500 randomly selected records, 758 (50.5%) were included. We retrieved or calculated 758 *P*-values for first reported outcomes and 389 for primary outcomes (specified in only 51% of included studies). The first reported and the primary outcome differed in 28% (110/389) of the included studies. The distributions of *P*-values for first reported outcomes and primary outcomes showed a notable discontinuity at the common thresholds of statistical significance (*P*-value = 0.05 and *P*-value = 0.01). We also found marked discontinuities in the distributions of z-scores across various medical disciplines, study designs, and types.

Conclusion: Reporting biases are still common in medical research. We discuss their implications, strategies to detect them, and recommended practices to avoid them.

Keywords: Bias; p-curve; p-hacking; Methodology; Reporting bias; Publication bias

1. Introduction

Complete publication of study results is essential to allow health care professionals and policy makers to make informed decisions. However, selective or distorted reporting is frequent in medical research [1]. Reporting biases arise if dissemination of research findings is influenced by the nature of the results. If undetected, reporting biases can lead to inaccurate conclusions and inappropriate decisions about health care and resource allocation, with potentially serious implications [2]. Failure to publish research findings honestly is unethical and a form of research misconduct [3,4]. Furthermore, research inaccessibility leads to waste of limited resources, unnecessary duplication, and loss of trust in scientific integrity [5].

Reporting biases may impact scientific reports in different ways [6–9]. First, a whole study may be suppressed, or harder to find, or published with delay, if its results are not considered to be interesting. The label "publication bias" is typically used to refer to this phenomenon [10]. Publication bias is the form of reporting bias that has been most extensively discussed in the literature over the last 60 years [11–13]. Second, results within a report of a study may be biased if the authors report the most interesting findings. For example, they may report the finding with smallest *P*-value or largest effect estimate after performing several analyses on the same outcome. Several terms have been coined to refer to such practice, including selective analysis reporting, data dredging, and p-hacking [14]. Alternatively, some outcomes that were measured and analyzed may be missing if the authors did not consider the results to be interesting.

Although these reporting biases are likely to have been always present in the dissemination of research findings,

Conflict of interest: The authors declare that there is no conflict of interest.

Funding: This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

* Corresponding author. Tel./fax: +61 (7) 5595 5201.

E-mail address: lnb6des@daad-alumni.de (L.N. Albarqouni).

http://dx.doi.org/10.1016/j.jclinepi.2016.11.013

0895-4356/

What is new?

Key findings

- There are discontinuities in the distributions of *P*-values from medical research studies at the typical thresholds of statistical significance that may provide indirect insights on reporting bias.
- Similar results were observed across various study designs and types.

What this study adds to what was known?

- Notable peaks in the distributions at common thresholds of statistical significant are consistent with either suppression of nonstatistically significant results or ''manipulation'' of reported findings to reach statistical significance.
- The outcome that is reported earliest in an article is more prone to this phenomenon than the primary outcome.

What is the implication and what should change now?

- The present investigation underpins the importance of the efforts and initiatives to tackle the mechanisms causing reporting biases (e.g., registration of studies, protocols, and statistical analysis).
- Researchers should continue to be encouraged to emphasize confidence intervals and effect sizes, rather than *P*-values, in the interpretation of results.
- There is a need for advocating the importance of replication, as well as the benefits of complete publication of research findings to reduce the prevalence of reporting biases in scientific literature.

more attention has been drawn to them recently due to the widespread use of systematic reviews. The validity of conclusions drawn from systematic reviews, intended to summarize the state of the art in a scientific area, is threatened if published results are not representative of the population of all conducted studies and analyses. Meta-analysis provides researchers with several graphical methods and statistical tests to assess the possible presence of reporting biases [6,10,13,15]. The exponential growth of published meta-analyses, many of them including some assessment of reporting biases, is likely to have increased the concern of incomplete publication of results as an ubiquitous problem in the scientific literature [8].

Evidence of reporting biases can be direct or indirect. Direct evidence includes tracking of cohorts of registered studies or conference proceeding abstracts and comparing the results of published and unpublished findings. For instance, studies have provided empirical evidence that studies with significant or positive results were more likely to be published, or more likely to be published earlier, than those with nonsignificant or unimportant results [5,8]. Direct evidence may also come from the acknowledgment of bias by those involved in the publication process, such as researchers, referees, and editors [16].

Indirect sources of evidence of reporting biases include the observation of a disproportionately high percentage of statistically significant findings in the published literature, as well as notable discontinuities in the *P*-value distribution curve just above the main significance thresholds ($P = 0.05$). Several papers have been published illustrating similar approaches in psychology, sociology, and natural science [14,17–19]. Here, we aim to explore indirect evidence of reporting biases by examining the empirical distribution of *P*-values reported in a large set of medical research studies and to compare this distribution across different contexts.

2. Methods

2.1. Study eligibility and selection

We conducted a descriptive cross-sectional survey of peer-reviewed, published, medical research articles. We sought original, primary, and quantitative research articles and searched the PubMed database using a simple search strategy that would identify most of these (Appendix A at www.jclinepi.com). We restricted the search to articles published in March 2014 and selected a random sample of 1,500 of the identified articles. To be included in the analyses, articles had to be written in English and had to involve only human participants. Articles had to include inferential statistics that investigated the efficacy or side effects of a medical or surgical intervention or investigating risk factors, exposures, or prognostic factors (epidemiological associations). We considered a wide range of study designs including randomized clinical trials, controlled clinical trials, before-after trials, cohort studies, case–control studies, and cross-sectional studies, and we considered a wide range of estimates including differences in means, risk ratios, odds ratios, hazard ratios, correlations, and regression coefficients. We included only articles that either reported the *P*-value or provided sufficient information to calculate a *P*-value for either the first reported or the primary outcome. We excluded duplicate reports of the same study as well as inaccessible full-text articles (e.g., published abstracts without full articles or study protocols).

2.2. Data screening and extraction

We developed a standardized data extraction form, which was pilot tested by all members of the research team. We extracted data based on the first reported outcome in the