

PROTOCOL

Open Access



# *P* value and Bayesian analysis in randomized-controlled trials in child health research published over 10 years, 2007 to 2017: a methodological review protocol

Alex Aregbesola<sup>1,2\*</sup>, Allison Gates<sup>3</sup>, Amanda Coyle<sup>1</sup>, Shannon Sim<sup>3</sup>, Ben Vandermeer<sup>3</sup>, Megan Skakum<sup>1</sup>, Despina Contopoulos-Ioannidis<sup>4</sup>, Anna Heath<sup>5,6</sup>, Lisa Hartling<sup>3</sup> and Terry P. Klassen<sup>1,2</sup>

## Abstract

**Background:** There is an unresolved debate about the reliability of the interpretation of *P* value. Some investigators have suggested that an alternative Bayesian method is preferred in conducting health research. As randomized-controlled trials (RCTs) are important in generating research evidence, we decided to investigate the extent, if any, the inferential statistical framework in published RCTs in child health research have changed over 10 years. We aim to examine the change in *P* value and Bayesian analysis in RCTs in child health research papers published from 2007 to 2017.

**Methods:** We will search the Cochrane Central Register of Controlled Trials (Wiley) to identify relevant citations. We will leverage a pre-existing sample of child health RCTs published in 2007 ( $n=300$ ) used in our previous study of reporting quality of pediatric RCTs. Using the same strategy and study selection methods, we will identify a comparable random sample of child health RCTs published in 2017 ( $n=300$ ). Eligible studies will include RCTs in health research among individuals aged 21 years and below. One reviewer will select studies for inclusion and extract the data and another reviewer will verify these. Disagreements will be resolved by a discussion between reviewers or by involving another reviewer. We will perform a descriptive analysis of 2007 and 2017 samples and analyze the results using both the frequentist and Bayesian methods. We will present specific characteristics of the clinical trials from 2007 and 2017 in tabular and graphical forms. We will report the difference in the proportion of *P* value and Bayesian analysis between 2007 and 2017 to assess the 10-year change. Clustering around *P* values of significance, if observed, will be reported.

(Continued on next page)

\* Correspondence: [alex.aregbesola@umanitoba.ca](mailto:alex.aregbesola@umanitoba.ca)

<sup>1</sup>The Children's Hospital Research Institute of Manitoba, 513-715 McDermot Avenue, John Buhler Research Centre, Winnipeg, Manitoba R3E 3P4, Canada

<sup>2</sup>Department of Pediatrics and Child Health, Max Rady College of Medicine, Rady Faculty of Health Sciences, University of Manitoba, Winnipeg, MB, Canada

Full list of author information is available at the end of the article



© The Author(s). 2021 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

(Continued from previous page)

**Discussion:** This review will present the difference in the proportion of trials that reported on  $P$  value and Bayesian analysis between 2007 and 2017 to assess the 10-year change. The implications for future clinical research will be discussed and this research work will be published in a peer-reviewed journal. This review has the potential to help inform the need for a change in the methodological approach from the null hypothesis significance test to Bayesian methods.

**Systematic review registration:** Open Science Framework <https://osf.io/aj2df>

**Keywords:**  $P$  value, Frequentist, Bayesian, Analysis, Child health, Randomized-controlled trials

## Background

Authors have continued to debate the reliance on  $P$  value alone in reporting and interpreting health research findings [1]. Chavalarias et al.'s [2] study from the USA examined the trend of  $P$  values and other statistical information reported in the entire MEDLINE database on biomedical research for over 25 years and found an increase in the reporting of  $P$  values over time. They also found that smaller  $P$  values were reported in the abstracts compared to the full-text, and the Bayesian methods were almost completely absent in the studies. Goodman et al.'s [3] report, also from the USA, which explored the properties and consequences of using Bayesian factors, found that the Bayes factor provides information about effect size and considers the alternative hypotheses of data compared to  $P$  value, which is computed with only the null hypothesis.

Accumulating studies have relentlessly highlighted the limitations and misconceptions of  $P$  values [4, 5]. One of such numerous misconceptions is the interpretation of a non-statistically significant difference ( $P$  value  $>5\%$ ) between two groups to mean that the null effect is most likely. This just means, however, that the null effect is statistically consistent with the observed results, including the range of effects in the confidence interval (CI) [4]. Likewise, equating statistical significance to clinical importance is erroneous because the difference may be too small to be clinically relevant. Sometimes, clinically relevant findings may not be statistically significant. While the use of  $P$  values may have a strong statistical history, compelling evidence showed that there is a need for complementary measures of evidence like effect sizes or replacing it with other inferential statistics such as Bayesian methods [6].

A study from Australia, which compared reporting research results with either the null hypothesis significance test (NHST, which is dependent on the  $P$  value) or confidence intervals (CIs), concluded that CIs elicit better interpretations if NHST is not invoked [7].

Some studies have also suggested that the subjective and arbitrary elements of  $P$  values are better clarified by Bayesian methods, which provide a more attractive alternative for better clinical trials [8]. A review that

compared frequentist NHST with Bayesian statistics in health research concluded that NHST is susceptible to confident misinterpretation, while Bayesian methods provide direct answers to how confident we should be in our results [9]. In an attempt to limit or eradicate misinterpretations associated with frequentist statistics, some studies have called for a complete ban of  $P$  values and NHST [10].

Following unresolved debate about the reliability of  $P$  value interpretation and the increasing interest in Bayesian methods [8], we decided to investigate the extent, if any, the inferential statistical framework in child health research has changed over 10 years [11]. We aim to examine the change in  $P$  value and Bayesian analysis and clustering around  $P$  values of significance in randomized-controlled trials (RCTs) in child health research papers published from 2007 to 2017.

## Methods

The present protocol has been registered within the Open Science Framework (registration: <https://osf.io/aj2df>) and is being reported in accordance with the reporting guidance provided in the Preferred Reporting Items for Systematic Reviews and Meta-Analyses Protocols (PRISMA-P) statement [12] (see checklist in Additional file 1). Any amendments made to this protocol when conducting the study will be outlined in the Open Science Framework and reported in the final manuscript

## Search strategy and study selection

We will leverage a pre-existing sample of child health RCTs published in 2007 ( $n = 300$ ) [11] used by our team in previous study of reporting quality of pediatric RCTs to answer our review question: What is the magnitude and direction of change in  $P$  value and Bayesian analysis reported in RCTs in child health research published over 10 years, if any? Details of the search strategy and study selection methods for the sample are available in our previous publications [11, 13]. We will replicate these methods to identify a comparable sample of child health RCTs published in 2017. The final sample will include 600 child-health RCTs, 300 published in each of 2007 and 2017.

To identify a sample of studies published in 2017, a research librarian will execute an updated literature search in the Cochrane Central Register of Controlled Trials (see Additional file 2). The Cochrane Central Register of Controlled Trials includes randomized and quasi-randomized controlled trials indexed in MEDLINE and EMBASE, hand-searched results, gray literature sources, and Cochrane Review Groups Specialized Registers of trials [14]. All retrieved records will be imported into EndNote (v. X9, Clarivate Analytics, Philadelphia, PA, USA) and exported to an Excel (v. 2016, Microsoft Corporation, Redmond, WA, USA) workbook for screening. We will randomly order the citations using the random numbers generator in Excel. Next, one reviewer will screen the titles and abstracts to identify the first 300 child health RCTs. These should be easily identifiable by title and abstract; however, in the unlikely (per experience) event that a record is deemed ineligible during data extraction, we will substitute it with the next relevant record. We will include the first 300 eligible citations from the randomly ordered list to make the sample size consistent with the previous publications [11, 12].

Eligible studies will include RCTs in health research conducted among individuals aged 21 years and below [15]. We will employ identical selection criteria used in the 2007 and 2012 samples to maintain consistency and comparability with earlier findings [11]. Literature will be limited to published full-text articles in the English language. There will be no restriction on settings in which the study was conducted, intervention, comparator, or the type of outcome.

#### Data extraction

We will adopt part of the data extraction form from the 2007 and 2012 studies [11], with some additions to gain the information on *P* values and Bayesian analysis. We will pilot test the form using three studies from 2007 and 2017 for completeness and accuracy. Data will be extracted by a single reviewer using Excel (v. 2016, Microsoft Corporation, Redmond, WA, USA), with verification by a second reviewer. Disagreements will be resolved by discussion between reviewers or by involving another reviewer when necessary. We will extract data on characteristics of the publication, study design, intervention, control, trial conduct, study sample, sample size, hypothesis, primary objective, diagnostic criteria, recruitment strategies, funding, data monitoring committee (DMC), and specific statistical attributes of frequentist and Bayesian analysis/methods that are related to the primary outcome (see Additional file 3). We will extract data for the primary outcome, and if not clearly stated, we will use the objective outcome (e.g., mortality, hospitalization), the outcome used to calculate sample size, or the first outcome reported in the results.

We will also use trial registers and published protocols (when cited in the publication) to supplement data extraction. When not cited in the publications, we will search for trial registers in the International Clinical Trials Registry Platform and the Google databases. We will not appraise the risk of bias of the included studies.

#### Data analysis

We will present summary characteristics and results of all trials in a tabular form. We will consider analyzing the data using Stata (v. 16.1; StataCorp, College Station, Texas, United States) or R [16] and JAGS statistical software [17]. The analysis of extracted data will be mainly descriptive, using counts and percentages for categorical data, and means and/or medians (with standard deviations and/or ranges) for continuous data. We will compare extracted data from the 2017 sample with 300 RCTs published in 2007 to assess 10-year change in the reporting of *P* value and Bayesian analysis. The difference between the two periods will be assessed using both the frequentist and Bayesian methods. We will present the proportion (%) of studies reporting *P* value and Bayesian analysis in 2007 and 2017 in graphical forms. We will also present specific characteristics of studies, which used Bayesian analysis in tabular forms, if any. We will present the clustering around *P* values of significance, if observed in the samples.

#### Discussion

To the best of our knowledge, this will be the first review to investigate the change in *P* value and Bayesian analysis in RCTs in child health research. This review will provide data on the methodological quality of RCTs in child health research, especially in the magnitude and direction of change in *P* value and Bayesian analysis in the 600 RCTs to be included in this review. Our experience with the two previous reviews will provide adequate guidance for study selection, data extraction, and interpretation of the results. We anticipate a considerable variation in the use of NHST and Bayesian methods in the 300 RCTs. Although the search strategy was clearly defined, we anticipate some limitations due to our inclusion criteria. Relevant studies may be omitted if not indexed in the databases we searched, full-text not available, or if reported in other languages other than English.

In conclusion, this review will provide robust evidence on the state of inferential statistics in RCTs in child health research. It has the potential to help inform which methodological approach should be adopted between NHST and Bayesian methods in RCTs in child health research.

## Study dissemination

We will submit reports from this study for peer-reviewed publication in appropriate academic journals. Our findings will be presented at provincial, national, and international scientific conferences and webinars. We will also share our findings via our institutional Twitter accounts.

## Abbreviations

NHST: Null hypothesis significance test; RCT: Randomized-controlled trials

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13643-021-01622-8>.

**Additional file 1.** PRISMA – P 2015 Checklist

**Additional file 2.** Search Strategy

**Additional file 3.** Data Extraction Guidelines

## Acknowledgements

The authors would like to thank Dr. Michele Dyson for her contribution to the 2007 sample used in this study. We also want to thank the administrative staff of the Children's Hospital Research Institute of Manitoba. LH is supported by a Canada Research Chair (Tier 1) in Knowledge Synthesis and Translation.

## Authors' contributions

AA and TPK contributed to the study conceptualization. AA, TPK, AG, and LH contributed to the study methods. AA drafted the protocol. AG and the research librarian designed the search strategy. AA, AG, AC, MS, and SS will be involved in study screening and data extraction. All authors read and approved final protocol.

## Funding

The Children Hospital Foundation of Manitoba

## Availability of data and materials

Not applicable

## Declarations

### Ethics approval and consent to participate

Not applicable

### Consent for publication

Not applicable

### Competing interests

The authors declare that they have no competing interests.

### Author details

<sup>1</sup>The Children's Hospital Research Institute of Manitoba, 513-715 McDermot Avenue, John Buhler Research Centre, Winnipeg, Manitoba R3E 3P4, Canada.

<sup>2</sup>Department of Pediatrics and Child Health, Max Rady College of Medicine, Rady Faculty of Health Sciences, University of Manitoba, Winnipeg, MB, Canada. <sup>3</sup>Department of Pediatrics and the Alberta Research Centre for Health Evidence (ARCHE), University of Alberta, Edmonton, Canada.

<sup>4</sup>Department of Pediatrics, Division of Infectious Diseases, Stanford University School of Medicine and Meta Research Innovation Center at Stanford (METR ICS), Stanford, CA, USA. <sup>5</sup>The Hospital for Sick Children and the University of Toronto, Toronto, ON, Canada. <sup>6</sup>University College London, London, UK.

Received: 18 September 2020 Accepted: 26 February 2021

Published online: 10 March 2021

## References

- Goodman SN. Toward evidence-based medical statistics: the P value fallacy. *Ann Intern Med.* 1999;130:995–1004.
- Chavalarias D, Wallach JD, Li AH, Ioannidis JP. Evolution of reporting P values in the biomedical literature, 1990–2015. *JAMA.* 2016;315:1141–8.
- Goodman SN. Toward evidence-based medical statistics. 2: The Bayes factor. *Ann Intern Med.* 1999;130:1005–13.
- Goodman S. A dirty dozen: twelve p-value misconceptions. *Semin Hematol.* 2008;45:135–40.
- Gelman A. P values and statistical practice. *Epidemiology.* 2013;24:69–72.
- Wetzels R, Matzke D, Lee MD, Rouder JN, Iverson GJ, Wagenmakers EJ. Statistical evidence in experimental psychology: an empirical comparison using 855 t tests. *Perspect Psychol Sci.* 2011;6:291–8.
- Coulson M, Healey M, Fidler F, Cumming G. Confidence intervals permit, but do not guarantee, better inference than statistical significance testing. *Front Psychol.* 2010;1:26.
- Lee JJ, Chu CT. Bayesian clinical trials in action. *Stat Med.* 2012;31:2955–72.
- Buchinsky FJ, Chadha NK. To P or not to P: backing Bayesian statistics. *Otolaryngol Head Neck Surg.* 2017;157:915–8.
- David Trafimow & Michael Marks. Editorial. *Basic and Applied Social Psychology.* 2015;doi: 10.1080/01973533.2015.1012991.
- Hamm MP, Hartling L, Milne A, Tjosvold L, Vandermeer B, Thomson D, et al. A descriptive analysis of a representative sample of pediatric randomized controlled trials published in 2007. *BMC Pediatr.* 2010;10:96.
- Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Syst Rev.* 2015;4:1-4053-4-1.
- Gates A, Hartling L, Vandermeer B, Caldwell P, Contopoulos-Ioannidis DG, Curtis S, et al. The conduct and reporting of child health research: an analysis of randomized controlled trials published in 2012 and evaluation of change over 5 years. *J Pediatr.* 2018;193:237–44.
- Cochrane Library [Internet]. Cochrane central register of controlled trials (CENTRAL). Hoboken: Wiley. <http://www.cochranelibrary.com/about/central-landing-page.html>. Accessed 27 Oct 2019.
- Hardin AP, Hackell JM. Committee on practice and ambulatory medicine. Age limit of pediatrics. *Pediatrics.* 2017;140:10.
- R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2019. <https://www.R-project.org/>
- Plummer, M. JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. In *Proceedings of the 3rd international workshop on distributed statistical computing.* 2003;124:1-10.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

