P-value and Bayesian analysis in randomized-controlled trials in child health research published in 2007 and 2017: a methodological review

Alex Aregbesola (Alex.Aregbesola@umanitoba.ca)
The Children's Hospital Research Institute of Manitoba
Allison Gates
University of Alberta
Amanda Coyle
The Children's Hospital Research Institute of Manitoba
Shannon Sim
University of Alberta
Ben Vandermeer
University of Alberta
Megan Skakum
The Children's Hospital Research Institute of Manitoba
Despina Contopoulos-Ioannidis
Department of Pediatrics, Stanford University School of Medicine and Meta Research Innovation Center at Stanford (METRICS), Stanford, CA
Anna Heath
The Hospital for Sick Children, University of Toronto
Lisa Hartling
University of Alberta
Terry P. Klassen
The Children's Hospital Research Institute of Manitoba

Research Article

Keywords: P-value, Bayesian methods, RCTs, child health, methodological review

Posted Date: June 13th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1722905/v1

Abstract

Background

Reliance on P-value of significance in clinical trials is a source of debate because of misconceptions and misinterpretations associated with it. Bayesian methods are suggested as an alternative approach. As randomized-controlled trials (RCTs) are essential in generating research evidence, we investigated the change in the use of *P*-values and Bayesian analysis and the clustering of *P*-values at key significance levels in child health RCTs published in 2007 and 2017.

Methods

We searched Cochrane Central Register of Controlled Trials to identify random samples of child health RCTs published in 2007 (n = 300) and 2017 (n = 300). Data on trial characteristics and analytic approaches were extracted. We analyzed the 600 RCTs using the frequentist and Bayesian methods. The change in the proportion of trials reporting P-values and Bayesian analyses was assessed using Pearson/Fisher Exact tests and non-informative Dirichlet priors.

Results

Of 600 RCTs, 535 (89%) used frequentist methods only versus 65 (11%) that included some Bayesian methods. Only 2 of the 65 trials used Bayesian inferential statistics. The use of frequentist methods decreased from (273, 91% to 262, 87%) while the inclusion of Bayesian analysis slightly increased from (27, 9% to 38, 13%) between 2007 and 2017. Although most RCTs were from Europe (172, 29%) and North America (133, 22%), the increase in proportion of trials by continent was most in Asia (mean difference (MD) = 0.14, 95% credible interval (CI) 0.08–0.20) with posterior probability *(PP*) of 1.00. Parallel (487, 81.2%) and cluster (58, 9.7%) RCTs were the most common RCT types but the increase in cluster RCT (0.06, 95%CI 0.02–0.11, *PP* = 0.99) was more than any other RCT types over 10 years. We found clustering of *P*-values at the significance level of 0.05 (437, 72.8%), which increased between 2007 (209, 69.7%) and 2017 (228, 76%). The smallest *P*-value reported in this review was 0.0001 (1, 0.2%).

Conclusions

The statistical framework in child health RCTs has not changed from the frequentist methods that is based on P-values with an unexplained clustering at the significance level of 0.05. Bayesian methods may increase the confidence in interpretation of results of RCTs

Background

A growing body of evidence calls for the use of Bayesian methods in conducting and analyzing randomized controlled trials (RCTs) [1, 2]. RCTs provide high-level evidence for clinical practice, so efforts are required to optimize the credibility of their results. A review that compared the frequentist null hypothesis significance testing (NHST) framework with Bayesian statistical methods in health research concluded that NHST is susceptible to misinterpretation [3]. In contrast, Bayesian methods provide direct answers to how confident we should be in our results [3]. Accumulating studies have relentlessly highlighted the limitations and misconceptions of the NHST framework and *P*-values [4, 5]. One of the 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 actually means, however, that the null effect is statistically consistent with the observed results, including the range of effects in the confidence interval (CI) [5]. Likewise, equating statistical significance to clinical significance is erroneous because the difference may be too small to be clinically relevant. Sometimes, clinically relevant findings may not be statistically significant. It is against the backdrop of these misinterpretations that authors have clamored for less confusing methods [2, 4].

A methodological review by Chavalarias et al. [6] from the United States 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, with Bayesian methods almost completely absent from these studies. Likewise, Goodman et al. [7] explored the properties and consequences of using Bayes factors as the inferential framework. They found that the Bayes factor provides information about effect size and considers the alternative hypotheses of data. This contrasts with *P*-values, which only consider the null hypothesis in their computation. Some other studies have suggested that Bayesian methods can help clarify the subjective and arbitrary elements of *P*-values and as such are more suitable for RCTs [8]. To limit or eradicate misinterpretations associated with frequentist statistics, some authors have argued that *P*-values and NHST should be abolished [9].

Following the call for more Bayesian methods in RCTs [1, 2, 8], we investigated the extent, if any, to which the inferential statistical framework in child health research has changed over 10-years [10]. We examined the change in the use of *P*-values and Bayesian analysis, and clustering of *P*-values at specific significance levels, in child health RCTs published in 2007 and 2017.

Methods

We registered the protocol within the Open Science Framework platform (registration ID: https://osf.io/aj2df). No amendments were made to the protocol while conducting the study. Our review question is: What is the magnitude and direction of change in the use of *P*-values and Bayesian analysis methods in child health RCTs published in 2007 and 2017, if any. We reported this review in accordance with the reporting guidance provided in the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) [11]. 1.

Study eligibility criteria

Eligible studies included RCTs in health research conducted among individuals aged 21 years and below published in 2007 and 2017 [12]. We employed identical selection criteria used in the 2007 sample to maintain consistency and comparability with earlier findings [10]. Our final samples were limited to full-text articles published in English language. We placed no restrictions on the settings in which the study was conducted, intervention, comparator or outcome.

Search strategy

We leveraged a pre-existing sample of child health RCTs published in 2007 (n = 300) [10], used by our team in a previous study of reporting quality of pediatric RCTs. Details of the search strategy and study selection methods for the 2007 sample are available in our previous publications [10, 13]. To identify a sample of studies published in 2017, a research librarian executed an updated literature search in the Cochrane Central Register of Controlled Trials (Additional file 1). 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].

Study selection

All retrieved records were imported into EndNote (v. X9, Clarivate Analytics, Philadelphia, PA, United States) and exported to an Excel (v. 2016, Microsoft Corporation, Redmond, WA, United States) workbook for screening. Consistent with the methods used to identify the 2007 sample [10, 13], we randomly order the citations using the random numbers generator in Excel. Next, one reviewer (A.A., A.G., A.C., S.S., or M.S.) screened the titles and abstracts to identify the first 300 child health RCTs and when a record was deemed ineligible during data extraction, we substituted it with the next relevant record. We included the first 300 eligible citations from the randomly ordered list [10, 13]. The final sample included 600 child-health RCTs, 300 published in each of 2007 and 2017 (Figure 1).

Data extraction

We adopted part of the data extraction form from the 2007 study [11], with some additions to gain the information on *P* values and Bayesian analysis methods. We pilot tested the form using three studies from 2007 and 2017 for completeness and accuracy. One reviewer (A.A., A.G., A.C., S.S., or M.S.) extracted the data using Excel (v. 2016, Microsoft Corporation, Redmond, WA, United States); a second reviewer verified the extractions. Disagreements between reviewers, which occurred in <2% of the studies, were resolved by discussion between reviewers. We extracted 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 were related to the primary outcome (Additional file 2). We extracted data for the primary outcome, and when this was not clearly stated, we used the objective outcome (e.g., mortality, hospitalization), the outcome used to calculate sample size, or the first outcome reported in the results. We used trial registers and published protocols

(when cited in the publication) to supplement data extraction. When not cited in the publications, we searched for trial registers in the International Clinical Trials Registry Platform and Google databases.

Data analysis

We performed frequentist analyses using Stata (v. 16.1; StataCorp, College Station, Texas, United States) and Bayesian analysis using the jags program, called from within R statistical software [15, 16]. The analysis was mainly descriptive, using counts and percentages to compare the characteristics of trials between 2007 and 2017. We assessed the change in the proportion of trials that reported the *P*-value and Bayesian analysis using Pearson/Fisher Exact tests. Multinomial distributions with non-informative Dirichlet priors were used for the Bayesian analysis [17]. A trial was described to have used Bayesian methods if any of the Bayesian inferential statistics or characteristics was used either in the methods (including hypothesis testing and analysis) or in the results of the study (Additional file 2). We performed a descriptive analysis to examine clustering of the *P*-value at specific significance levels and presented it in a graph. We investigated the predictors of using any element of Bayesian analysis using a logistic regression analysis.

Results

The difference in the characteristics of trials published in 2007 and 2017 using the frequentist and Bayesian analysis is presented in Table 1 and Table 2, respectively. Four of the trial characteristics examined were significantly associated with the publication year (Table 1). Most of the RCTs were conducted in Europe (172, 28.7%) and North America (133, 22.1%), while South America (21, 3.5%) had the smallest number of trials. Parallel (487, 81.2%) and cluster (58, 9.7%) RCTs were more commonly used than the other RCT types. The proportion of trials that were conducted in a single center (309, 51.5%) was higher than those conducted across multiple centers (244, 40.7%). Likewise, the proportion of trials that did not report their minimal clinically important difference (MCID, 551, 91.8%) was higher than those that reported the MCID (49, 8.2%).

Table 1
Characteristics of child health RCTs published in 2007 and 2017 as assessed by the frequentist methods

Trial characteristics		**Publication year**		***P-value***
	Total n (%)	2007 n (%)	2017 n (%)	
Continent				< 0.001
Africa	46 (7.7)	18 (6)	28 (9.3)	
Asia	121 (20.2)	39 (13.0)	82 (27.3)	
Australasia	36 (6.0)	17 (5.7)	19 (6.3)	
Europe	172 (28.7)	105 (35.0)	67 (22.3)	
Multi-continent	47 (7.8)	23 (7.7)	24 (3.0)	
NR	24 (4.0)	17 (5.7)	7 (2.3)	
North America	133 (22.1)	71 (23.7)	62 (20.7)	
South America	21 (3.5)	10 (3.3)	11 (3.7)	
RCT type				0.013
Cluster	58 (9.7)	20 (6.7)	38 (12.7)	
Crossover	33 (5.5)	21 (7.0)	12 (4.0)	
Factorial	8 (1.3)	7 (2.3)	1 (0.3)	
Other (specify)	7 (1.2)	5 (1.7)	2 (0.7)	
Parallel	487 (81.2)	244 (81.3)	243 (81.0)	
Split body	7 (1.2)	3 (1.0)	4 (1.3)	
Control type				0.088
Active Intervention	265 (44.2)	138 (46.0)	127 (42.3)	
No intervention	77 (12.8)	32 (10.7)	45 (15.0)	
Other (specify)	99 (16.5)	46 (15.3)	53 (17.7)	
Placebo	127 (21.2)	72 (24.0)	55 (18.3)	
Usual care	1 (0.2)	1 (0.3)	0 (0.0)	
Wait-list control	31 (5.2)	11 (3.7)	20 (6.7)	
Number of centers				< 0.001

Trial characteristics		Publication year		*P-value*
Multicenter	244 (40.7)	112 (37.3)	132 (44.0)	
Single center	309 (51.5)	142 (47.3)	167 (55.7)	
Unclear	47 (7.8)	46 (15.3)	1 (0.3)	
Study hypothesis				0.204
A priori/alternative	70 (11.7)	33 (11.0)	37 (12.3)	
NR	335 (55.8)	180 (60.0)	155 (51.7)	
Null hypothesis	36 (6.0)	16 (5.3)	20 (6.7)	
Null hypothesis + a priori	1 (0.2)	1 (0.3)	0 (0.0)	
Other (specify)	158 (26.3)	70 (23.3)	88 (29.3)	

Table 1
cont. Characteristics of child health RCTs published in 2007 and 2017 as assessed by the frequentist methods

Trial characteristics		**Publication year**		***P-value***
	Total n (%)	2007 n (%)	2017 n (%)	
Power of trial calculated				0.278
No	555 (92.5)	281 (93.7)	274 (91.3)	
Yes	45 (7.5)	19 (6.3)	26 (8.7)	
Sample size calculated				0.378
No	309 (51.5)	176 (58.7)	133 (44.3)	
Yes	291 (48.5)	124 (41.3)	167 (55.7)	
Interim analysis reported				0.261
No	570 (95.0)	282 (94.0)	288 (96.0)	
Yes	30 (5.0)	18 (6.0)	12 (4.0)	
Primary outcome analysis based on				0.487
95% confidence interval	41 (6.8)	21 (7.0)	20 (6.7)	
P-value + Bayesian inferential	2 (0.3)	0 (0.0)	2 (0.7)	
P-value only (frequentist)	410 (68.3)	210 (70.0)	200 (66.7)	
Other (specify)	147 (24.5)	69 (23.0)	78 (26.0)	
MCID reported				0.003
No	551 (91.8)	266 (88.7)	285 (95.0)	
Yes	49 (8.2)	35 (11.7)	14 (4.7)	

NR, not reported; RCT, randomized-controlled trial; other, any type not covered in the specified categories, and these can be found in the extraction guidelines; MCID, minimal clinically important difference

Table 2
Characteristics of child health RCTs published in 2007 and 2017 as assessed by the Bayesian methods

Trial characteristics	Mean difference between 2007 and 2017			
	Mean difference	95% CI lower	95% CI upper	Probability difference > = 0
Continent				
Africa	0.033	-0.010	0.076	0.937
Asia	0.141	0.077	0.204	1.000
Australasia	0.007	-0.033	0.043	0.634
Europe	-0.125	-0.196	-0.054	0.000
Multi-continent	0.003	-0.040	0.046	0.560
NR	-0.030	-0.096	0.036	0.188
North America	-0.033	-0.066	-0.002	0.018
South America	0.003	-0.028	0.033	0.580
RCT type				
Cluster	0.060	0.015	0.108	0.994
Crossover	-0.030	-0.067	0.007	0.054
Factorial	-0.020	-0.040	-0.002	0.013
Other (specify)	-0.010	-0.030	0.007	0.123
Parallel	-0.003	-0.066	0.060	0.466
Split body	0.003	-0.014	0.022	0.648
Control type				
Active Intervention	-0.036	-0.118	0.043	0.189
No intervention	0.043	-0.010	0.097	0.942
Other (specify)	0.023	-0.038	0.081	0.775
Placebo	-0.056	-0.122	0.009	0.045
Usual care	-0.003	-0.014	0.005	0.182
Wait-list control	0.030	-0.004	0.066	0.950
Number of centers				
Multicenter	0.066	-0.013	0.141	0.951

Trial characteristics	Mean difference between 2007 and 2017			
Single center	0.083	0.002	0.159	0.982
Unclear	-0.149	-0.191	-0.110	0.000
Study hypothesis				
A priori/alternative	-0.019	-0.124	0.085	0.361
Null hypothesis	0.004	-0.076	0.087	0.544
Null hypothesis + a priori	-0.009	-0.035	0.010	0.151
Other (specify)	0.024	-0.093	0.142	0.065
Power of trial calculated				
Yes	0.012	-0.010	0.033	0.859
Sample size calculated				
Yes	0.071	0.025	0.120	0.998

Table 2
cont. Characteristics of child health RCTs published in 2007 and 2017 as assessed by the Bayesian methods

Trial characteristics	**Mean difference between 2007 and 2017**			
	Mean difference	95% CI lower	95% CI upper	Probability difference > = 0
Power of trial calculated				
Yes	0.012	-0.010	0.033	0.859
Sample size calculated				
Yes	0.071	0.025	0.120	0.998
Interim analysis reported				
Yes	-0.010	-0.028	0.008	0.134
Primary outcome analysis based on				
95% confidence interval	-0.003	-0.043	0.037	0.443
P-value + any Bayesian	0.007	-0.004	0.019	0.926
P-value only (frequentist)	-0.033	-0.109	0.039	0.186
Other (specify)	0.030	-0.038	0.097	0.805
MCID reported				
Yes	-0.033	-0.056	-0.012	0.002

RCTs, randomized-controlled trial; CI, credible interval; NR, not reported; other, any type not covered in the specified categories, and these can be found in the extraction guidelines

From the Bayesian analysis, there was an increase in the proportion of trials conducted in Asia (mean difference (MD) = 0.14, 95% credible interval (CI) 0.08–0.20) and Africa (0.03, 95%CI − 0.01–0.08). The posterior probabilities (*PP*) that these differences represent a true increase in the proportion of trials were equal to 1.0 and 0.94 for Asia and Africa, respectively (Table 2). The increase in proportion of cluster RCTs (0.06, 95%CI 0.02–0.11, *PP* = 0.99) was larger than for any of the other RCT types in the 600 trials. The reporting on the number of trial centers improves, meaning that the proportion of single-centered trials (0.083, 95%CI 0.002–0.159, *PP* = 0.98) and multi-centered trials (0.066, 95%CI − 0.013–0.141, *PP* = 0.95) both increased over the 10-year period. In contrast, the reporting of MCID slightly decreased (− 0.03, 95%CI − 0.056-–0.012, *PP* = 0.002). The proportion of RCTs that calculated power of trial (MD = 0.01, 95%CI − 0.01–0.03, *PP* = 0.86) or sample size (MD = 0.07, 95%CI 0.03–0.12, *PP* = 0.99) had a minimal increase. None of the trial characteristics was associated with using Bayesian methods in the trial reporting and analysis (Additional file 3; Table 3 and Additional file 4; Table 4).

Of the 600 RCTs, 535 (89%) trials used frequentist only versus 65 (11%) that used some Bayesian analysis in their report (Fig. 2). Between 2007 and 2017, the use of frequentist methods decreased (273, 91% versus 262, 87%), and the inclusion of Bayesian analysis slightly increased (27, 9% versus 38, 13%). Only 2 of the 65 trials used Bayesian inferential statistics in their analysis. We observed that *P*-values clustered at the commonly selected significance level of 0.05 (437, 72.8%) (Fig. 3), which increased between 2007 (209, 69.7%) and 2017 (228, 76%). The smallest reported *P*-value in this review was 0.0001 (1, 0.2%).

Discussion

We found that most of the studies in our randomly selected samples of child RCTs used a frequentist approach and that Bayesian methods are still uncommon. Around one-tenth of trials used both frequentist method and some elements of Bayesian methods in conducting the RCT. Only two studies used Bayesian inferential statistics in their analysis, and no study adopted the Bayesian approach entirely from study design through to the analysis and interpretation of results of the trial. The difference in the proportion of trials by selected characteristics between 2007 and 2017, as assessed by the frequentist and Bayesian approaches, was similar. The Bayesian methods provided additional information that allowed determining the key drivers of the differences observed over 10 years. We also found clustering of the *P*-values at the commonly selected significance level of 0.05, which increased between 2007 and 2017.

Reliance on P-value in the frequentist approach to clinical trials has continued to be a source of debate because of common misconceptions and misinterpretations that can lead to difficulties applying the results of research [18, 19]. As RCTs are essential in generating evidence for practice, there is a need to explore other approaches that may serve as an alternative or, at the very least, complement the frequentist methods when conducting RCTs. While we observed a statistically significant difference in a few of the selected characteristics of RCTs between 2007 and 2017 using the frequentist approach, the Bayesian analysis, provided more detailed information. In one of the trial characteristics examined, the frequentist analysis showed a statistically significant association between continent and publication year, but the Bayesian analysis showed the size of the increase or decrease observed between these two years for each continent, with the corresponding posterior CI and the probability that this increase or decrease was greater than 0. We showed from the Bayesian analysis that the probability of the increase observed in the proportion of trials from Africa and Asia occurring was strong. There was some evidence that the proportion of trials has also decreased in North America. In another characteristic, the frequentist analysis showed a statistically significant association between RCT type and publication year. However, the Bayesian analysis showed that this difference was due to a large increase observed in the proportion of cluster RCTs occurring over 10 years. We observed similar findings with other trial characteristics when examined using both inferential frameworks.

Bayesian clinical trials appear to be gaining momentum among clinical researchers in recent years [20]. The robustness and adaptive properties [3, 8] of Bayesian methods, which improve confidence in

interpreting trial results [21], may be partly responsible for the recent rise in interest. However, it is unclear whether Bayesian methods are a *better* alternative to the frequentist. A previous Canadian methodological review that re-analyzed 88 published RCTs using Bayesian statistical inference found that introducing Bayesian *PP* (with a well-defined prior) offered a better interpretation of RCTs results. They further reported that the *PP* of the recorded large benefits in the RCTs with positive findings (obtained using the frequentist approach) were lower and variable [22]. Their findings are in keeping with our results in which we tested the difference in trial characteristics of 600 RCTs using both frequentist and Bayesian approaches. We found that the *PPs* of the difference occurring during the 10 years studied were varied and not completely consistent with the strength of the association observed via the frequentist approach as indicated by the *P*-values. Another relatively recent Swedish study reanalyzed a specific RCT (designed to investigate the effect of electronic screening and brief intervention on harmful alcohol consumption of Swedish students) and found no statistically significant association between the intervention and control groups ($P = 0.13$) [23]. However, a Bayesian reanalysis of these data indicated that there was *PP* of 0.93 that the intervention group consumed less alcohol versus the control.

In keeping with Chavalarias et al. [6]'s study from the United States that found almost no Bayesian methods in a collection of published articles from MEDLINE examined over 25 years, we also found only 2 articles that adopted Bayesian inferential statistics. These findings suggest that there might be some *P*-hacking or publication bias towards reporting trials or clinical research with statistical significance [24]. Bayesian methods are not based on specifying a threshold for the *P*-value that indicates statistical significance or not and do not aim to reduce the risk of false positive conclusions from a trial, as is used in the frequentist approach [25]. Instead, they combine prior information and the data collected in the study to make probability statements about the likely values of the quantity of interest, e.g., odds ratio. It is beyond the scope of this study to determine the reason that the frequentist approach is dominant in RCTs. Nevertheless, it is essential to reiterate that the risk of chasing statistical significance in RCTs is palpable, and it is time we focused on reducing this effect by considering Bayesian methods as part of the possible solution.

Strengths And Limitations

This study is not without strengths and limitations. Our experience with the two previous reviews [10, 13] provided adequate guidance for study selection, data extraction and interpretation of the results. To the best of our knowledge, this is the first methodological review to show clustering of the *P*-value at the significance level of 0.05 in child health RCTs. The analysis in this review were conducted using frequentist and Bayesian methods that allowed opportunity for comparison of interpretation of results based on both approaches. We had a large pool of samples of Child RCTs so that if a particular study did not meet our inclusion criteria, we were able to substitute with other studies, which supported the timely completion of this study. While our sample did not contain all the child health RCTs published during the studied periods, we performed a random sampling of RCTs with no restrictions on the settings in which the study was conducted, intervention, comparator or outcome to reflect the landscape of all child health RCTs. We did not assess the risk of bias, which could have provided an extra layer of information on the

reporting quality of the studies that adopted some elements of Bayesian characteristics within the 600 trials.

Future Research

The next steps should involve looking into the use of Bayesian analysis in recently published RCTs in child health research to confirm or refute the suggestions from this review. Since some of the concerns about frequentist methods are related to the analysis and, more importantly, the misinterpretation of results from the analysis, more methodological studies should look at re-analyzing results of RCTs conducted by the frequentist methods.

Conclusions

In conclusion, the predominant inferential statistical framework in child health RCTs has not changed from the historical NHST, and there is evidence of substantial clustering of *P*-values at the commonly used threshold of significance of 0.05. The analysis of this review suggests that Bayesian analysis may provide an additional layer of confidence in the interpretation of result of child health RCTs.

Abbreviations

CI
Confidence interval
CI
Credible interval
DMC
Data Monitoring Committee
MCID
Minimal clinically important difference
MD
Mean difference
NHST
Null hypothesis significance testing
NR
Not reported
RCT
Randomized-controlled trial
PP
Posterior probability
PRISMA
Preferred Reporting Items for Systematic Reviews and Meta-Analyses

Declarations

Acknowledgements

The authors thank Clara Tam for assistance in the design of the tables in the manuscript.

Authors' contributions

AA, AG, LH, and TPK conceived the idea and designed the study. AA, AG, AC, MS, SS, developed and conducted the search, screened the studies and performed data extraction. AA and AH analyzed and interpreted the data. AA wrote the first draft. All authors read and approved the final manuscript for publication.

Funding

The Children's Hospital Foundation of Manitoba supported this study

Availability of data and materials

All data generated and analysed during this study are included in this published article.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare they have no competing interests.

References

1. Biard L, Bergeron A, Levy V, Chevret S. Bayesian survival analysis for early detection of treatment effects in phase 3 clinical trials. Contemp Clin Trials Commun 2021; 21:100709.
2. Bendtsen M. A Gentle Introduction to the Comparison Between Null Hypothesis Testing and Bayesian Analysis: Reanalysis of Two Randomized Controlled Trials. J Med Internet Res 2018; 20:10873.
3. Buchinsky FJ, Chadha NK. To P or Not to P: Backing Bayesian Statistics. Otolaryngol Head Neck Surg 2017; 157:915–18.
4. Goodman S. A dirty dozen: twelve p-value misconceptions. Semin Hematol 2008; 45:135–140.
5. Gelman A. P values and statistical practice. Epidemiology 2013; 24:69–72.

6. Chavalarias D, Wallach JD, Li AH, Ioannidis JP. Evolution of Reporting P Values in the Biomedical Literature, 1990–2015. JAMA 2016 Mar 15; 315:1141–1148.
7. Goodman SN. Toward evidence-based medical statistics. 2: The Bayes factor. Ann Intern Med 1999; 130:1005–1013.
8. Lee JJ, Chu CT. Bayesian clinical trials in action. Stat Med 2012; 31:2955–2972.
9. Trafimow D, Marks M. Editorial. Basic and Applied Social Psychology 2015; 37:1–2
10. 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; 10:96-2431-10-96.
11. Page MJ, McKenzie JE, Bossuyt PM, *et al.* The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021; 372
12. Hardin AP, Hackell JM, COMMITTEE ON PRACTICE AND AMBULATORY MEDICINE. Age Limit of Pediatrics. Pediatrics 2017; 140:2017–2151.
13. 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; 193:237–244.e37.
14. Cochrane Library. Cochrane central register of controlled trials (CENTRAL)., https://www.cochranelibrary.com/central/about-central. Accessed October 27, 2019.
15. R Core Team. A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www R-project org/ 2019.
16. Plummer, M., 2003, March. JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. In Proceedings of the 3rd international workshop on distributed statistical computing (Vol. 124, No. 125.10, pp. 1–10).
17. Gelman A, Jakulin A, Pittau M, Su Y. A weakly informative default prior distribution for logistic and other regression models. The Annals of Applied Statistics. 2008. 2:1360–1383
18. Goodman SN. Toward evidence-based medical statistics. 1: The P value fallacy. Ann Intern Med 1999; 130:995–1004.
19. Sterne JA, Davey Smith G. Sifting the evidence-what's wrong with significance tests? BMJ 2001; 322:226–231.
20. Lee JJ, Yin G. Principles and Reporting of Bayesian Trials. J Thorac Oncol 2021;16:30–36.
21. Bautista JR, Pavlakis A, Rajagopal A. Bayesian analysis of randomized controlled trials. Int J Eat Disord 2018;51:637–646.
22. Wijeysundera DN, Austin PC, Hux JE, Beattie WS, Laupacis A. Bayesian statistical inference enhances the interpretation of contemporary randomized controlled trials. J Clin Epidemiol 2009; 62:13–21.e5.
23. Bendtsen M. An Electronic Screening and Brief Intervention for Hazardous and Harmful Drinking Among Swedish University Students: Reanalysis of Findings From a Randomized Controlled Trial

Using a Bayesian Framework. J Med Internet Res 2019; 21:e14420.
24. The ASA Statement on p-Values: Context, Process, and Purpose. The American Statistician. ISSN: 0003-1305, 1537–2731.
25. Smith GD, Ebrahim S. Data dredging, bias, or confounding. BMJ 2002; 325:1437–1438.
26. ; NR, not reported (in cases where the statistical significance was not reported in the paper)

Figures

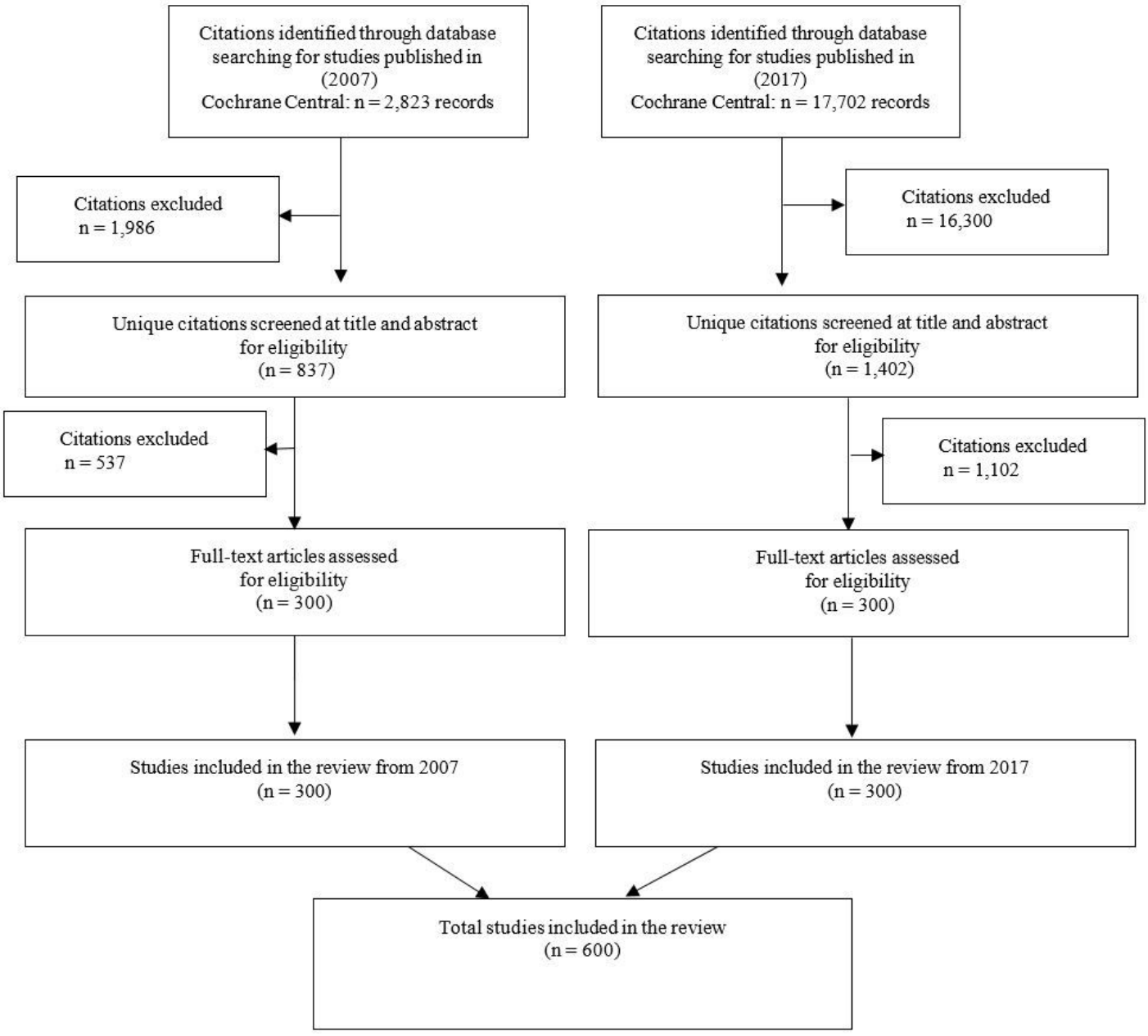

Figure 1

Summary of literature search and screening process

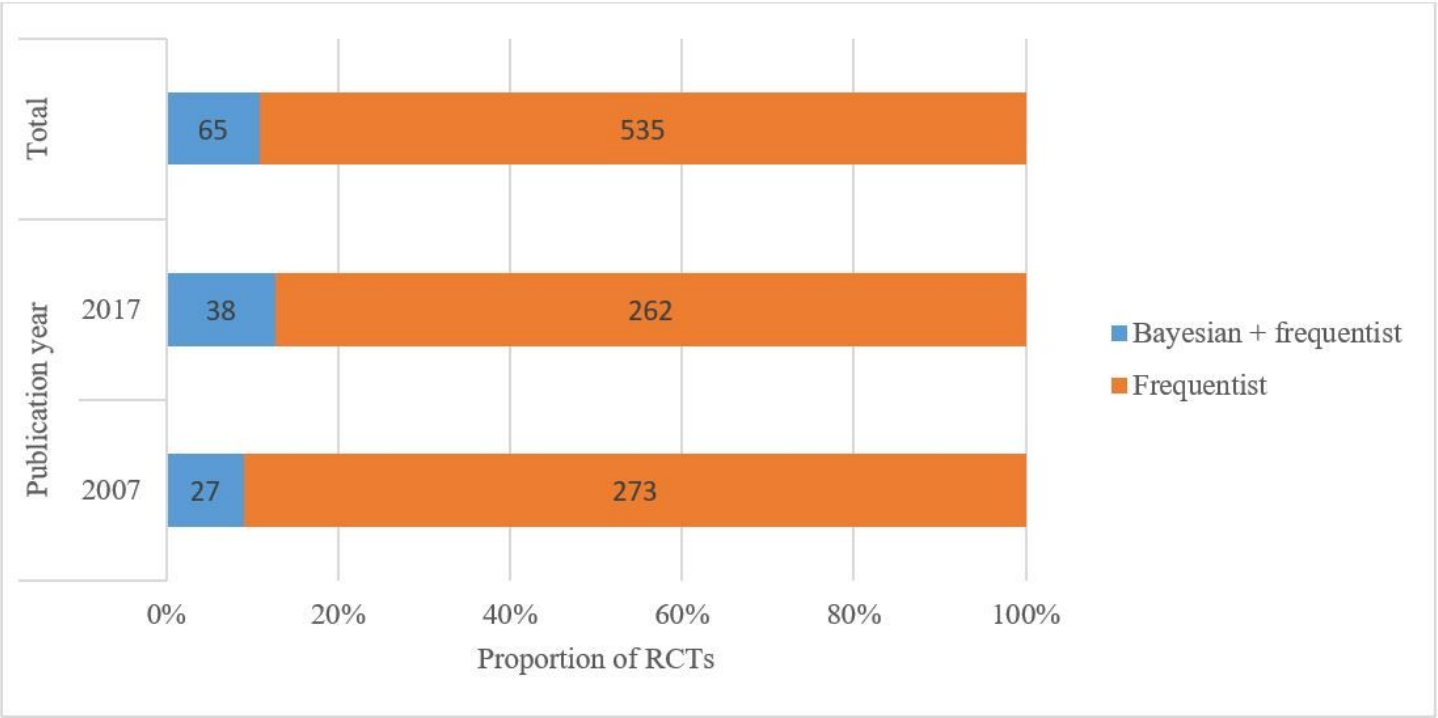

Figure 2

Proportion (%) of trials that used Bayesian versus frequentist methods in child health randomized-controlled trials (RCTs) published in 2007 and 2017

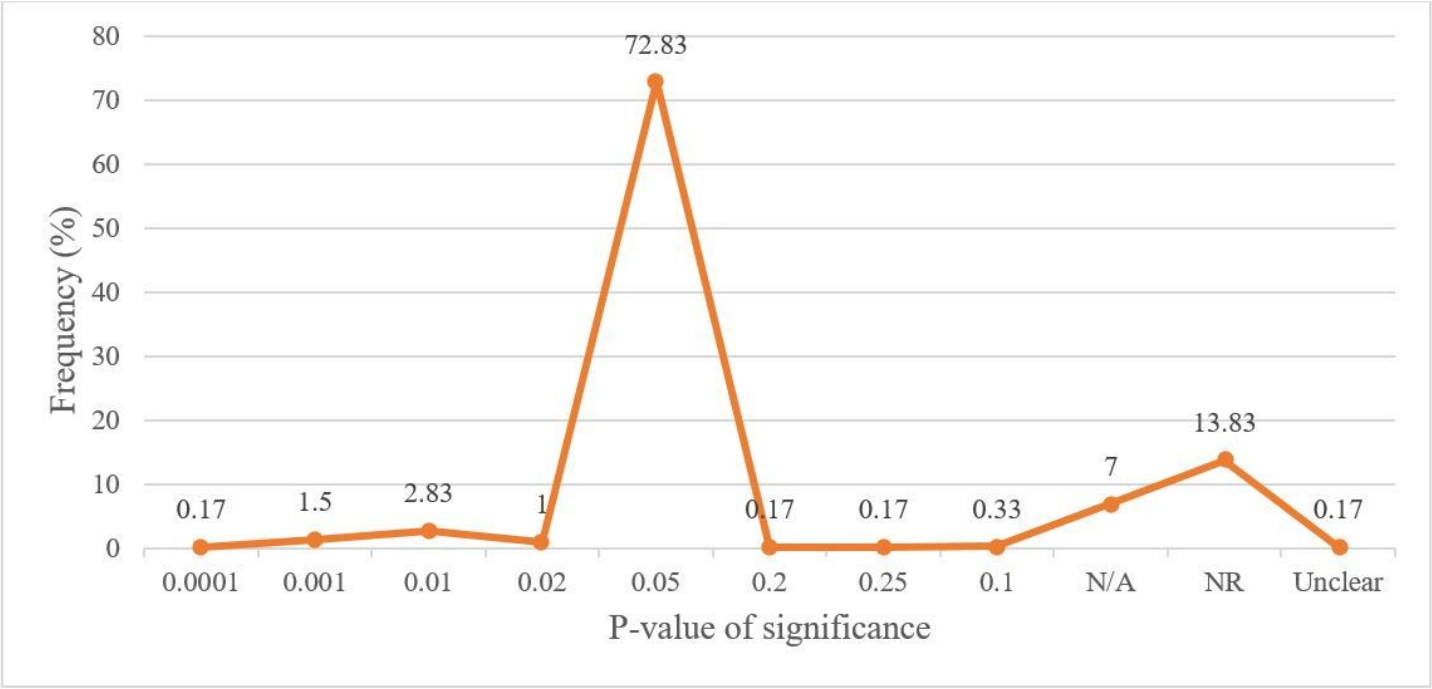

Figure 3

Clustering of *P*-values at specific significance levels in child health RCTs published in 2007 and 2017. RCT, randomized controlled-trials; N/A, not applicable (in cases where the statistical significance reported was not related to the primary outcome; NR, not reported (in cases where the statistical significance was not reported in the paper)

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Additionalfile1Searchstrategy.docx
- Additionalfile2Dataextractionguidelines.docx
- AdditionalfileResultsTable3.docx
- AdditionalfileResultsTable4.docx