BJA

British Journal of Anaesthesia, 125 (2): 208–215 (2020)

doi: 10.1016/j.bja.2020.04.093
Advance Access Publication Date: 20 June 2020
Special Article

A consensus checklist to help clinicians interpret clinical trial results analysed by Bayesian methods

David Ferreira[1,2,*], Mael Barthoulot[3], Julien Pottecher[4], Klaus D. Torp[5], Pierre Diemunsch[4] and Nicolas Meyer[2,6]

[1]Anesthesiology and Intensive Care Department, CHU Besançon, Besançon, France, [2]Université de Strasbourg, iCUBE, UMR7357, Illkirch Cedex, France, [3]Institut Pasteur de Lille, Unité d'Epidémiologie et de Santé Publique, Inserm-U1167, Lille, France, [4]Anesthesiology and Intensive Care Department, CHU de Strasbourg, IHU-Strasbourg, Strasbourg, France, [5]Department of Anesthesiology, Mayo Clinic, Jacksonville, FL, USA and [6]Public Health Department, CHU de Strasbourg, GMRC, Strasbourg, France

*Corresponding author. E-mail: dferreira@chu-besancon.fr

Abstract

Introduction: In the context of an increasing number of publications of trial data analysed by Bayesian methods, clinicians need support to better understand Bayesian statistical methods. The existing checklists are intended for people who already know these methods. We aimed to establish and validate a checklist that contains a group of items considered crucial in interpreting the results of a phase III RCT analysed with Bayesian methods.
Methods: A team of biostatisticians created a checklist of previously reported items and additional items identified from a literature review. Using three different articles in three rounds, the items were then validated by residents in anaesthesiology with no skills in statistics.
Results: Based on an initial item list, three rounds led to a consensus checklist. Eleven items were considered important information to be specified for understanding the validity of the results. Of these, three were considered essential: specification of the prior, source of the prior (when prior is informative), and the effect size point estimate with its credible interval.
Conclusion: The checklist can help clinicians interpret the results of a phase III randomised clinical trial analysed by Bayesian methods, even clinicians with no particular knowledge of statistics, to ensure that the major elements of the statistical section are present and valid. Care should be taken in interpreting the results of a trial analysed by Bayesian methods that are not reported with these three essential items because the validity of the results cannot be established.

Keywords: Bayesian; biostatistics; checklist; clinical trial; RCT; reproducibility; statistical methods

Editor's key points

- Checklists can assist readers to understand and interpret clinical trials, in a manner akin to clinical checklists.
- Bayesian statistical methods are increasingly used for statistical analysis of clinical trials.
- The authors constructed a checklist of items required for understanding the validity of results derived by Bayesian methods.
- This checklist can help clinicians interpret the results from clinical trials that have been analysed using Bayesian methods.

Received: 12 August 2019 **Accepted:** 6 April 2020

Clinicians usually read innovative literature to resolve daily clinical problems. Checking the validity of the methodology of a study is crucial and depends highly on the statistical aspects of the study. In this context, many guidelines have been created to help report the findings, such as CONSORT (Consolidated Standards of Reporting Trials), STROBE (Strengthening The Reporting of OBservational Studies in Epidemiology), PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses), or STARD (STAndards for the Reporting of Diagnostic accuracy studies).[1–4] Among the different parts of an article, the 'Statistical analysis' subsection (under Methods) is both particularly important and difficult to grasp for many readers, most mainly having a clinical background. In a local unpublished survey, we observed that most clinicians did not read the statistical part of a report of an RCT because it was considered too difficult to understand. Clinicians also felt that they were not adequately trained to critically read this section or just did not feel concerned by this part.

The mainstream statistical method is the so-called classical or frequentist method. The past 2 decades, with the improvement in computers, has witnessed renewed interest in an older statistical concept, Bayesian statistics.[5,6] Many papers have demonstrated the wide applicability and value of the methods in clinical trials.[7–11] To facilitate Bayesian reporting, some guidelines were implemented, such as BayesWatch, BaSiS (Bayesian Standards in Science), or ROBUST (Reporting Of Bayes Used in Clinical Studies).[12–14] However, these guidelines are difficult for clinicians to use because they are not meant to help in interpretation and comprehension by non-statistician readers. These works were more for guidance in the writing and reporting of articles rather than reading trial articles and therefore were intended for an informed audience only. The users of such tools are supposed to be already acquainted with Bayesian statistics. Nevertheless, despite the increase in the number of publications of trial data analysed by Bayesian methods,[15] clinicians are still poorly trained in Bayesian methods. Publications are usually written by a multidisciplinary team consisting of clinicians, biologists, and statisticians; few are written by authors who master each part of the study and its manuscript. The same holds true for readers of a publication, who may also lack the required background.

Thus, in terms of statistical analysis, new tools are needed to help clinicians identify the essential items in research articles so that they can critically evaluate the statistical methods used, even if they do not completely understand the methods used. Without a thorough understanding of what is implied by Bayesian methods, the clinician could still search for items that may help determine whether the reporting of results of a Bayesian trial is appropriate and whether the article reports the essential elements of the statistical analysis to be valid.

The aim of this study was to create and validate a checklist to identify the crucial items of the statistical part of Bayesian publications. This checklist is intended to be used by readers without any knowledge of Bayesian methods (clinicians, editors, reviewers, and investigators). Moreover, its purpose is not to ensure that readers will thoroughly understand every detail of the Bayesian statistics, although we hope that they will gain a better understanding of these methods. For this secondary goal, the general concepts of the Bayesian analysis are described in an accompanying publication.[16] To illustrate and clarify each item of our Bayesian interpretation checklist, we refer to a recently published trial with results analysed by Bayesian methods, the IMMERSION trial.[17]

Methods

The checklist was created and validated in six steps.

Step 1. Collection of items from previous guidelines

Guidelines for reporting Bayesian methods were searched. Each item of the identified reading grids was included and clarified to be understandable to a clinician with no knowledge of biostatistics. The item list was generated by a team of 7 biostatisticians who were specialised in the use of Bayesian methods (Strasbourg University hospital methodological unit).

Step 2. Identification of other items or sub-items from a literature review

We aimed to identify items other than those identified in the first step that might be included. Items were searched from the statistical sections of identified reports of phase III RCTs.[15]

Step 3. Formulation of checklist keywords

For each item identified in Step 1 or 2, we searched reports of trials included in the review for terms that may match the item domain and thus be considered keywords. The aim of this step was to improve the correct identification rate of each item, despite a possibly limited knowledge of Bayesian statistical terminology. The characteristic of each item was classified as 'comprehension' or 'reproducibility'. 'Reproducibility' refers to the possibility for an author to replicate the study in a subsequent similar investigation using the related information. When an item was related to the study reproducibility, its sole presence was considered relevant. When the item was missing, the possibility of replicating the study must be critically evaluated. 'Comprehension' refers to the fact that a missing item may be a source of bias or omission, and the reader must search for this bias.[16]

Step 4. Selection and reformulation of items by the working group

We established an experimental version of the tool with the pre-selected items. This version was examined by the working group to reach consensus. The objective of this step was to select a small set of items and reformulate them so that clinicians can check for their presence in a manuscript. Then, considering that in a reading grid, not all items on the list had the same weight, members of the working group were asked to identify those that are essential for validating the results of a trial analysed by Bayesian methods.

Step 5. Reformulation and validation by clinicians

We recruited residents without knowledge of statistics and Bayesian methods to use the checklist as they read several reports of RCTs to evaluate the clarity of each items. The aim of this step was to evaluate whether a physician could check for the presence or absence of each matching keyword in the report without having to completely understand the concept behind each item. The expected result was the unexperienced reader being able to have a comprehensive view of the Bayesian methods described in the paper. The ease with which

each item was found was scored on a four-point scale: 1, very easy to identify; 2, easy to identify; 3, difficult to identify; 4, very difficult or impossible to identify. The item in the report was matched with the corresponding checklist item. After each trial, residents were asked to explain their scoring. If the score was 3 or 4, residents had to describe their understanding of the item. They were also asked to describe their understanding of the items they had not found. Descriptions of the items were then adjusted accordingly and residents re-tested the new modified checklist with the same methodology on a new randomly selected trial until all items had been rated 1 or 2 by all residents, and each item identified by residents matched the items identified by checklist authors. The checklist was then considered validated.

Step 6. Evaluation of the practical use of the reading grid

The validated reading grid was tested by clinicians in real practice using an online questionnaire. Volunteers with no training in statistics or Bayesian methods were provided with randomly selected RCT, and a list of keywords covering the different list items. They had to search for each item in each specific article part, by noting the time it took to find them.

Explanation and illustration of the checklist

We used the IMMERSION study[17] to illustrate our checklist. The IMMERSION trial used only Bayesian methods. The main objective of the IMMERSION study was to determine the effect of immersion on diuresis and haemodynamic variables in young women. It aimed at demonstrating a reduction in blood volume with an increase in hourly diuresis, after partial immersion in an obstetric dilatation bath. This was an open prospective interventional study of two randomised and controlled parallel groups.

Results

Step 1. Collection of items from previous guidelines

Three previously published guidelines can be used to help investigators report the Bayesian analysis of trial results: BayesWatch,[12] BaSiS,[13] and ROBUST.[14] These guidelines were intended for regular users of Bayesian methods considering that the wording required good knowledge of the Bayesian jargon. All items of these checklists were reformulated with simplified keywords for non-statistician readers, to help in identifying critical points of the Bayesian analysis.

Step 2. Identification of other items or sub-items

We identified reports of 49 phase III clinical trials that used Bayesian methods for analysis.[15] From the methods and results of these publications, two sub-items specific to phase III trials were identified in addition to those of the first step.

Step 3. Formulation of checklist keywords

The keywords are shown in Table 1.

Step 4. Selection and reformulation of items by a working group

An experimental version of the tool included 20 preselected items. The working group reached consensus, and seven items were considered to have limited interest because their validity could not be checked. The remaining items are shown in Table 1, and among these, three items were identified as crucial for validating the results of a trial analysed by Bayesian methods: specification of the prior, source of the prior (in the context of informative prior), and presentation of the point estimate and its credible interval.

Step 5. Reformulation and validation by clinicians

Reports of three clinical trials were used to validate the checklist.[18–20] Nine residents in anaesthesiology participated in the validation, with at least two participating in each round (Table 2). With the final article, each resident clearly identified all items (items rated '1' or '2'). Items were included in the checklist to facilitate comprehension and assessment of the reproducibility of the trial. The final checklist is presented in Table 1.

Step 6. Evaluation of the practical use of the reading grid

Forty-eight clinicians used the reading grid on an anaesthesia article.[21] Participating clinicians were from four different centres. There were 21 (43.8%) anaesthetists and intensive care specialists, 19 (39.6%) medical specialists, and 8 (16.6%) surgeons, 21 (43.8%) of whom were interns and 27 (56.2%) doctors. On average, they identified 11.1 out of 14 items (standard deviation [SD] 2.1) in 12.4 min (SD 5.7) (Table 2).

Explanation and illustration of the checklist

The following is a description of each item of the checklist. Each item is introduced in logical order of appearance in the paper. The section was specified to inform the clinicians where items must be searched. Each item was classified as 'comprehension' or 'reproducibility'.

In the methods section

Item 1. Objective

Was the question of interest written in a probabilistic language? The objective must be presented with wording such as 'what was the likelihood that the relative risk was larger than 1 or larger than 1.5' rather than as 'the objective is to estimate a relative risk of … ' This formulation expresses the more relevant question that investigators do indeed ask: What was the probability that the treatment was clinically relevant?

Example:

The objective of the IMMERSION trial, '*to estimate the probability [of] an increase in hourly diuresis after partial water immersion*', was written in a probabilistic language.

Model building

Item 2. Prior (or a priori distribution)

Item 2.1 Was the *a priori* distribution of the main outcome specified? (Distribution, parameters): for example normal N (μ,σ^2) or binomial (n,p) or any other distribution.

Table 1 Checklist for evaluating results of a randomised controlled trial analysed by Bayesian methods. Each item (and keywords) must be searched in the section where it logically belongs. The comprehension (C) and reproducibility (R) domain are specified for each item. Keywords are for helping the clinician check for the presence of items.

Part	Item	Description	Keywords	Characteristic	Reported on page
METHODS	**1. Objective**	Was the question of interest written in a probabilistic language?	Probability; likelihood	C	
Model building	**2. Prior** (or *a priori* distribution)	2.1 Was the *a priori* distribution of the main outcome specified?	Prior (noun not adverb); *a priori*; distribution; variable; normal; binomial	R	
		2.2 Was the prior informative?	Informative; inform; uninformative; sceptical; optimistic; pessimistic; enthusiastic; Jeffreys	C	
		2.3 Was the prior defined prospectively?	Prospective	C	
		2.4 What data or knowledge was the prior based on? (if the prior is informative)	Previous study; review; meta-analysis; RCT; elicitation; expert opinion	C	
	3. Model	Was the statistical model specified?	Model; regression; linear; logistic; Cox	R	
Model implementation	**4. Software**	Was the software specified?	r-project in the method or bibliographical part; Bugs; Fortran; mathematic; stata; winBUGS; openBUGS; JAGS; SAS	R	
	5. Settings	Were the settings of the algorithm specified?	Number of iterations; burn in; thinning; stationarity	R	
	6. Validity criteria	Did the algorithm converge?	Algorithm; converge	C	
RESULTS	**7. Data and likelihood**	Was a summary of the data for the experiment available?	Mean; standard deviations (SD); proportions; median; inter-quartile range (IQR)	R	
	8. Description of results	Was the point estimate of the effect size or the relevant parameter value and its credible interval shown?	Mean; probability; credible interval (CrI)	C	
	9. Posterior distribution	Was the posterior distribution shown?	Figure showing a plot of posterior distribution	C	
	10. Sensitivity analysis	Was the impact of different priors (optimistic or sceptical) explored?	Sensitivity analysis; prior modification; sceptical; enthusiastic; optimistic; pessimistic; Jeffreys	C	
DISCUSSION	**11. Advantages and limitations of Bayesian analysis**	Were the advantages and limitations of the Bayesian analysis presented?	Bayesian analysis; Bayesian analysis limitations; Bayesian analysis benefits; Bayesian analysis advantages	C	

Table 2 Description of the validation of the checklist items by clinicians. Number of readers who rated a checklist item as easy or very easy to identify ('1' or '2') in reports of results of RCTs analysed with Bayesian methods and who correctly copied that given item to the matching Bayesian checklist item for each round of evaluation. Practical use corresponds to the number of readers who correctly identified the items present or absent when using the reading grid in real practice.

	Checklist item														No. of readers
	1	2.1	2.2	2.3	2.4	3	4	5	6	7	8	9	10	11	
Round 1	3	1	1	1	1	5	5	4	0	0	3	0	0	2	5
Round 2	2	2	2	2	2	2	2	2	1	1	1	1	2	1	2
Round 3	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6
Practical use	37	33	32	40	24	41	38	43	44	40	35	42	43	41	48

The *prior* distribution corresponds to the available information on the variable of interest before data collection. Was the distribution appropriate for the data type? Were the selected parameters relevant? Was the prior relevant in the context?

Example:

In the IMMERSION trial, the prior distribution was specified: 'The different priors used in the analyses of the main outcome are listed in Table 2a in the 'Results' section. For the bath group, priors were N(1,0.25), N(1,1), N(1,10), N(2,0.25), N(2.68,1). For the bed group, priors were N(1,0.25), N(1,1), N(1,10), N(2,0.25), N(1.75,1.56)'.

Item 2.2 Was the prior informative? A prior distribution is considered informative if it expressed precise information about the parameter of interest, such as effect size. A non-informative or uninformative prior is considered vague and indicates high uncertainty on the parameter of interest.

Example:

In the IMMERSION trial, the prior distribution was informative: '*The priors were based on physiological knowledge*'.

Item 2.3 Was the prior defined prospectively? When had the prior been defined? Was the prior defined before the data collection?

Bias can be induced if the prior was developed after data collection, especially if it was based on expert opinion.

Example:

In the IMMERSION trial, the statistical analysis was developed prospectively. '*The priors were defined before the study*'.

Item 2.4 Which data or knowledge was the prior based on? Meta-analysis, previous study, expert opinion, elicitation etc.

It is important to specify the prior and its form. The validity of a prior derived from a trial or meta-analysis is less controversial than one derived from expert opinion, which appears more subjective. In this case, the authors must explain and justify the method of elicitation.

Example:

In the IMMERSION trial, the priors were based on physiology: '*The priors were defined before the study and were based on the basis of physiological knowledge*'.

Item 3. Model

Was the statistical model specified? For example linear regression, logistic regression.

Example:

In the IMMERSION trial, '*We computed the mean difference and its 95% credible [interval], and the probability that the difference [was] positive*'.

Model implementation

Item 4. Software

Was the software specified? For example: R, WinBUGS, OpenBUGS, and JAGS.

Example:

In the IMMERSION trial, the software was specified. '*All computations were done with R 3.2.2 and JAGS statistical software with all the required additional packages*'.

Item 5. Settings

Were the algorithm and the settings of the algorithm specified? The algorithm used for the estimation, Markov chain Monte Carlo (McMC); number of chains used and initial values.

Example:

In the IMMERSION trial, the details of the algorithm used for the estimation were specified: '*Convergence of the McMC sample chain [was] checked graphically*'.

The settings include the number of iterations (number of McMC simulations planned), number of iterations burned (number of McMC iterations before chain convergence) and thinning (process allowing to keep one simulation for every *X* simulations).

Example:

In the IMMERSION trial, the settings were specified: '*A burn in of 5000 iterations followed by 100 000 iterations was used for each analysis*'.

Item 6. Validity criteria

Did the algorithm converge? The algorithm convergence is required for the model to be valid.

Example:

In the IMMERSION trial, the McMC algorithm converges: '*Convergence of the McMC sample chain [was] checked graphically*'.

In the results section

Item 7. Data and likelihood

Was a summary of the data for the experiment available? (means, SD, proportions, median, inter-quartile range).

Example:

In the IMMERSION trial, 'Results are displayed in Table 2a' with means and 95% credible intervals.

Item 8. Description of results

Was the point estimate of the effect size or the relevant parameter value and its credible interval shown? A point estimate of the effect size and its credible interval ('range' for the parameter of interest) represents a summary of the posterior distribution that expresses all the information and knowledge available at the end of the study on the effect size measure.

Example:

In the IMMERSION trial, the point estimate and its credible interval were presented in Table 2a. 'For the difference [Bath-Bed], the mean difference [95% CrI] was 1.26 [0.20; 2.32]'.

Item 9. Posterior distribution

Was the posterior distribution shown? This involves a graphical representation of the main result of the study (posterior distribution).

Example:

In the IMMERSION trial, the posterior distribution of diuresis was shown in Figure 1.

Item 10. Sensitivity analysis

Was the impact of different priors (lowly informative or informative, enthusiastic, or sceptical) explored? What was the relative weight of the prior against data? Was a sensitivity analysis using another distribution performed? If the posterior distribution was only weakly modified and influenced, whether one uses or not any of several notably different prior distributions (i.e. favouring or not the expected result), it indicates a robust result, depending mainly on the data. Such a result was also susceptible to agreement among experts previously expressing diverging opinions.

Example:

In the IMMERSION trial, the sensitivity analysis was specified: 'Comparisons were performed under the Bayesian paradigm, taking into account priors, either neutral or based on the results from Katz *et al.*'.

In the discussion section

Item 11. Advantages and limitations of Bayesian analysis

Were the advantages and limitations of the Bayesian analysis presented?

Example:

In the IMMERSION trial, the advantages and limits of the analysis were not presented.

Discussion

We present a checklist to help clinicians critically read and interpret original reports of results of RCTs analysed with Bayesian methods. This checklist, based on expert knowledge, was validated by clinicians for clinicians. In the reading grids, items do not have the same weight. Among the seven items in the list, three were considered essential for interpreting results of RCTs analysed with Bayesian methods: specification of the prior, source of the prior (when prior is informative), and presentation of the point estimate and its credible interval. Priors can entirely modify results; thus, their description and justification are essential in reports of RCTs. A scientific article with results analysed by Bayesian methods that does not include these three items is of limited interest because the validity of the results cannot be checked. The absence of one of these three items should be considered as D-dimer in pulmonary embolism. Their presence does not prove that the paper has a perfect statistical methodology, but their absence makes it possible to rule out a good paper with a large negative predictive value. If we were provocative, we would say that the absence of these three items should cause the article to be rejected without even reading it because it is impossible to check its validity. Their presence does not guarantee the correct use of Bayesian methods, but it does allow verifications.

To validate our checklist, we used a sort of 'anti-DELPHI' method considering that the residents included were 'expert in the lack of knowledge' in Bayesian methods (i.e. perfectly inexperienced). Several rounds were necessary to clarify items for clinicians. The simplification of the description of items and identification of keywords were a crucial point to allow the clinician to identify them. Items such as settings of the prior distribution and its parameters may be of paramount

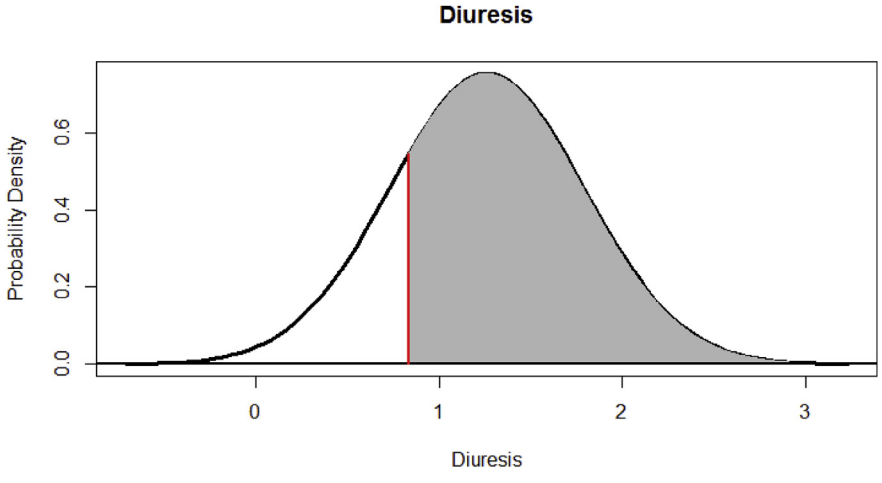

Fig 1. Diagram of the posterior distribution of diuresis in the IMMERSION trial.[17]

importance for statisticians, but a lay clinician may not be able to fully grasp the details of their specification, even with the help of a simplified checklist. During the step of validation of the checklist, the residents had to justify their answers and they spontaneously copied on their answer grids these items because they intuitively identified their importance. In the practical use of the reading grid, clinicians, from different centres, without any knowledge of Bayesian methods and for the first use of the reading grid, were able to identify in 12 min an average of 11 items out of 14. The least well-reported item was the justification of the prior because they had not looked for the keyword in the right section. In fact, the keyword for item 2.4 was present in the Introduction section (the phrase cited by the people who identified this keyword), but outside of this section it had no value, as understood by half the group who had followed the instructions completely. Moreover, this performance is quite good considering that the volunteers only had the keywords to find the items (Table 1). They did not have the content of the present work to help them find the items.

The use of our final checklist helped readers without experience in statistics and Bayesian methods to identify crucial items without the goal of complete comprehension of the methods. As with a 'before take-off' aeronautic pilot checklist, all security points must be checked. In clinical practice, before modifying practice according to new results from an RCT, the validity of the results, with the statistical analysis, must also be checked.[22] This kind of checklist could be a part of the solution to reduce misinterpretation or not reading the Statistical analysis section[23] resulting from methodological deficiencies of an article or a defect in training or other reasons (lack of time, not reading because of statistical technicalities etc.). This checklist may also help authors in drafting their research reports.

The limitation of the checklist is that clinicians do not have to understand the method if they apply it in a literal way as a list of items to check. Another limitation is that users cannot check implicit items. Each specialty has its own jargon, and statisticians are no exception. In a context of multidisciplinary publications, clarity must be the rule to enable each reader to find crucial items. The checklist will at least have the advantage of discriminating between good- and poor-quality papers. The interested clinician may consult additional material further explaining the Bayesian method.[16]

The presence in a given paper of all of the items in our checklist is not sufficient to guarantee that the study is statistically correct from all its aspects. To check this point, in-depth comprehension of the statistical part is required, and our checklist only allows for ensuring that the required items identified are duly presented in the article.

Conclusions

This article presents a checklist for evaluating the results of RCTs analysed by Bayesian methods that was validated by clinicians inexperienced in statistical analysis. The primary objective of this checklist is to help readers of a given article, even with no particular knowledge in statistics, use the Bayesian paradigm to ensure that major items in the statistics section are present. The checklist is necessary but not sufficient for determining the quality of a Bayesian trial, which explains why we have created another document designed to help the interested reader better understand the main concepts of Bayesian methods.[16] We consider that results involving the use of Bayesian analysis that do not describe three essential items in the statistics section (prior specified, source of the prior [when the prior is informative], presentation of the point estimate and its credible interval) should be interpreted with caution.

Authors' contributions

Study design: DF, MB, KT, PD, NM
Data collection: DF, JP
Manuscript preparation: DF, JP, PD, NM
Manuscript revision: DF, MB, JP, KT, PD, NM

Funding

Centre Hospitalier Universitaire de Strasbourg, France.

Declarations of interest

The authors declare that they have no conflicts of interest.

Presentation

Preliminary data for this study were presented as a poster presentation at the EPICLIN meeting, 9–12 June 2019, Toulouse, France, and will be presented at the CFIES meeting, 25–27 September 2019, Strasbourg, France, and at the ASA meeting, 19–23 October 2019, Orlando, FL, USA.

Acknowledgements

The authors thank the anaesthesia residents who validated the reading grid: R. Alardin, Y. Essaber, C. Jeanclerc, M. Ambs, E. Dupont, T. Grelet, C. Louyot, T. Vieille, P.A. Poinsot. Groupe Méthodes en Recherche Clinique (GMRC), Public Health Service, Strasbourg University Hospital: EA Sauleau, F Lefèvre, F Severac, M Schaeffer, J Muller. They also thank Laura Smales (BioMedEditing, Toronto, Canada) for English editing.

References

1. Altman DG, Schulz KF, Moher D, et al. The revised CONSORT statement for reporting randomized trials: explanation and elaboration. *Ann Intern Med* 2001; **134**: 663–94
2. Von Elm E, Altman DG, Egger M, Pocock SJ, Gotzsche PC, Vandenbroucke JP. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *BMJ* 2007; **335**: 806–8
3. Liberati A, Altman DG, Tetzlaff J, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. *BMJ* 2009; **339**: b2700
4. Bossuyt PM, Reitsma JB, Bruns DE, et al. The STARD statement for reporting studies of diagnostic accuracy: explanation and elaboration. *Ann Intern Med* 2003; **138**: W1–12
5. Gilks WR, Thomas A, Spiegelhalter DJ. A language and program for complex Bayesian modelling. *Statistician* 1994; **43**: 169
6. Lunn DJ, Thomas A, Best N, Spiegelhalter D. WinBUGS — a Bayesian modelling framework: concepts, structure, and extensibility. *Stat Comput* 2000; **10**: 325–37

7. Berry Statistics. *A bayesian perspective* 1996
8. Spiegelhalter DJ, Abrams KR, Myles JP. *Bayesian approaches to clinical trials and health-care evaluation*. New York: John Wiley & Sons; 2004
9. Goodman SN. Introduction to Bayesian methods: I. Measuring the strength of evidence. *Clin Trial*. 2005; **2**: 282–90. discussion 301–4, 364–78
10. Louis TA. Introduction to Bayesian methods II: fundamental concepts. *Clin Trial*. 2005; **2**: 291–4. discussion 301–4, 364–78
11. Berry DA. Introduction to Bayesian methods III: use and interpretation of Bayesian tools in design and analysis. *Clin Trial. Lond Engl* 2005; **2**: 295–300. discussion 301–304, 364–78
12. Spiegelhalter DJ, Myles JP, Jones DR, Abrams KR. Bayesian methods in health technology assessment: a review. *Health Technol Assess* 2000; **4**: 1–130
13. The BaSiS Group. *Bayesian standards in science (BaSiS) [Internet]*. Draft: Sept. 13 2001. http://lib.stat.cmu.edu/bayesworkshop/2001/BaSis.html
14. Sung L, Hayden J, Greenberg ML, Koren G, Feldman BM, Tomlinson GA. Seven items were identified for inclusion when reporting a Bayesian analysis of a clinical study. *J Clin Epidemiol* 2005; **58**: 261–8
15. Ferreira D, Vivot A, Diemunsch P, Meyer N. Bayesian analysis from phase III trials was underused and poorly reported: a systematic review. *J Clin Epidemiol* 2020; **123**: 107–13
16. Ferreira D, Barthoulot M, Pottecher J, Torp KD, Diemunsch P, Meyer N. Theory and practical use of Bayesian methods in interpreting clinical trial data: a narrative review. *Br J Anaesth* 2020; **125**: 201–7
17. Effect of Immersion. *Performed under the conditions of obstetrical dilatation bath*. on Diuresis and Hemodynamic Variables in Young Women — Full Text View — ClinicalTrials.gov [Internet]. [cited 2018 Jul 24]. Available from: https://clinicaltrials.gov/ct2/show/NCT02409953
18. Tinmouth A, Tannock IF, Crump M, et al. Low-dose prophylactic platelet transfusions in recipients of an autologous peripheral blood progenitor cell transplant and patients with acute leukemia: a randomized controlled trial with a sequential Bayesian design. *Transfusion (Paris)* 2004; **44**: 1711–9
19. Reddy VY, Sievert H, Halperin J, et al. Percutaneous left atrial appendage closure vs warfarin for atrial fibrillation: a randomized clinical trial. *JAMA* 2014; **312**: 1988–98
20. Sananes N, Roth GE, Aissi GA, et al. Acupuncture version of breech presentation: a randomized sham-controlled single-blinded trial. *Eur J Obstet Gynecol Reprod Biol* 2016; **204**: 24–30
21. Kranke P, Thompson JP, Dalby PL, et al. Comparison of vestipitant with ondansetron for the treatment of breakthrough postoperative nausea and vomiting after failed prophylaxis with ondansetron. *Br J Anaesth* 2015; **114**: 423–9
22. Baker M. 1,500 scientists lift the lid on reproducibility. *Nature* 2016; **533**: 452–4
23. McCloskey Z. The standard error of regressions. *J Econ Lit* 1996: 97–114

Handling editor: Gareth Ackland