The reporting quality and transparency of orthopaedic studies using Bayesian analysis requires improvement: A systematic review

Faris Bdair a, Sophia Mangala b, c, Imad Kashir c, Darren Young Shing c, John Price d, Murtaza Shoai e, Breanne Flood i, Samera Nademi f, Lehana Thabane b, c, Kim Madden c, g, *

a Mathematical and Computational Science, Stanford University, USA
b Department of Health Research Methods, Evidence & Impact, McMaster University, Canada
c Research Institute of St. Joseph’s Hamilton, Canada
d University of Southern California, USA
e Department of Molecular Biosciences, University of Kansas, USA
f Department of Medicine, McMaster University, Canada
g Department of Surgery, McMaster University, Canada

ARTICLE INFO

Keywords:
Bayesian analysis
Systematic review
Orthopaedic surgery

ABSTRACT

Background: Bayesian methods are being used more frequently in orthopaedics. To advance the use and transparent reporting of Bayesian studies, reporting guidelines have been recommended. There is currently little known about the use or applications of Bayesian analysis in orthopedics including adherence to recommended reporting guidelines. The objective is to investigate the reporting of Bayesian analysis in orthopedic surgery studies; specifically, to evaluate if these papers adhere to reporting guidelines.

Methods: We searched PUBMED to December 2nd, 2020. Two reviewers independently identified studies and full-text screening. We included studies that focused on one or more orthopaedic surgical interventions and used Bayesian methods.

Results: After full-text review, 100 articles were included. The most frequent study designs were meta-analysis or network meta-analysis (56%, 95% CI 46–65) and cohort studies (25%, 95% CI 18–34). Joint replacement was the most common subspecialty (33%, 95% CI 25–43). We found that studies infrequently reported key concepts in Bayesian analysis including, specifying the prior distribution (37–39%), justifying the prior distribution (18%), the sensitivity to different priors (7–8%), and the statistical model used (22%). In contrast, general methodological items on the checklists were largely well reported.

Conclusions: There is an opportunity to improve reporting quality and transparency of orthopaedic studies using Bayesian analysis by encouraging adherence to reporting guidelines such as ROBUST, JASP, and BayesWatch. There is an opportunity to better report prior distributions, sensitivity analyses, and the statistical models used.

1. Introduction

Bayesian statistical methods implement Bayes’s theorem, which states that the data moderate prior beliefs regarding the model parameters, and this process produces updated beliefs about model parameters [1]. This is important in medicine because it allows for a redistribution of credibility over different parameters in our data. Essentially, this makes it easy to more clearly see how our data affects any prior beliefs or hypotheses. When it comes to Bayesian analysis, results are favorable because they can be directly interpreted unlike in frequentist analysis, where results only tell us how unlikely our set of data is [2]. A systematic review of Bayesian articles in psychology revealed that the use of Bayesian methods in applied psychological work has steadily increased since the nineties [1] and the number of studies using a Bayesian approach continues to increase [3]. Several studies on reporting of Bayesian analyses in epidemiology research [3], phase III trials [4], and N-of-1 trials [5] suggest that the quality of reporting on Bayesian analyses is poor and could use improvement. There was, however, no literature pertaining to the reporting of Bayesian analyses in orthopedics.

* Corresponding author. Department of Surgery, Department of Health Research Methods, Evidence & Impact, McMaster University, G841-50 Charlton Ave E, Hamilton, Ontario, L8L 4A6, Canada.
E-mail address: maddenk@mcmaster.ca (K. Madden).

https://doi.org/10.1016/j.conctc.2023.101132
Received 24 March 2022; Received in revised form 20 March 2023; Accepted 5 April 2023
Available online 7 April 2023
2451-8654/© 2023 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).
There are several guidelines for reporting Bayesian analyses, each of which has many similar domains of interest, but has slightly differing wording or instructions. We have selected three commonly used Bayesian analysis reporting checklists that vary slightly in their approach. The Reporting Of Bayes Used in clinical Studies (ROBUST) guidelines were created in 2004 [6]. These guidelines consisted of seven items: the specification and justification of the prior distribution, sensitivity analysis, reporting of the statistical model, the analytical technique, central tendency, and variance [6]. The Jeffreys’s Amazing Statistics Program (JASP) guidelines were developed in 2020 with a twelve-item checklist [7]. JASP includes many similar items as the ROBUST checklist and adds items on stating the goal of the analysis, reporting the Bayes factor, graphing the prior and posterior distribution, reporting sequential analyses, and availability of statistical files online [7]. The BayesWatch checklist includes 12 items including statistical model, loss function, prior distribution, computations, posterior distributions, sensitivity analyses, and several general methodology items [8].

The research question is: among orthopaedic surgical studies that use Bayesian analyses, how well does the published literature adhere to the ROBUST, BayesWatch, and JASP reporting guidelines?

2. Methods

This is a systematic review of existing literature in orthopedic surgery that used Bayesian analyses.

2.1. Study inclusion

We searched Pubmed from inception to December 2nd, 2020 using the terms “(Orthopedic OR Orthopaedic) AND (Bayesian Analysis).” Two reviewers (FB and SM) independently identified studies for potential eligibility based on titles and abstracts using the Covidence systematic review software. Disagreements were resolved through conversation or by consulting a senior reviewer (KM). Two reviewers (FB and SM) independently (i.e. masked to each other’s rating) conducted full-text screening in duplicate using Covidence after title and abstract screening. We included studies with various objectives, such as therapy studies, diagnostic studies, economic evaluations, and prevalence/incidence studies. We excluded studies that did not involve patients undergoing orthopaedic surgical interventions. For example, we excluded studies of non-surgical treatments such as physiotherapy for knee arthritis. We included papers that were primary studies, meta-analyses, or systematic reviews. Opinion pieces and protocols were excluded, and we included only English language studies.

2.2. Data extraction

Two reviewers independently extracted data from each study into a study-specific data extraction form in the REDCap electronic data capture system. Three teams of two reviewers independently extracted the items on the ROBUST (FB and SM), BayesWatch (DYS and IK), and JASP (JP and MS) checklists. Two reviewers (BF and SN) also extracted study characteristics for each paper, including type of orthopedic surgery, statistical software and statistics used in the analysis, and key demographic characteristics. Discrepancies were resolved by a senior reviewer (KM).

2.3. Statistical Analysis

The analyses are descriptive. We present frequency data (proportions and percentages, and 95% CI where applicable) to describe the percentage of studies that fully reported checklist items for each of the different guidelines. We used Rstudio version April 1, 1106 for thee analyses.

3. Results

Our search in Pubmed identified 241 articles for potential inclusion (Fig. 1). After title and abstract screening, we excluded 113 study articles for the following reasons: not related to orthopedics (n = 71), did not involve a surgical treatment (n = 22), protocol/methodology paper only (n = 9), there was no abstract available (n = 4), design other than a primary study or systematic review (n = 4), and the study was not in humans (n = 3). After full-text review, we excluded a further 28 studies for the following reasons: did not involve a surgical treatment (n = 22), and there was no Bayesian analysis (n = 6). We finally included 100 articles (Fig. 1). Most studies were conducted in North America (38%, 95% CI 29–48), Asia (37%, 95% CI 28–47) and Europe (19%, 95% CI 13–28). The most common study designs were meta-analysis or network meta-analysis (56%, 95% CI 46–65) and cohort studies (25%, 95% CI 18–34). Joint replacement was the most common orthopaedic subspecialty (33%, 95% CI 25–43) (Table 1).

Fig. 1. Study flow diagram.
Table 1
Study characteristics.

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Number</th>
<th>n = 100</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Region</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>North America</td>
<td>38</td>
<td>29-48</td>
<td></td>
</tr>
<tr>
<td>Asia</td>
<td>37</td>
<td>28-47</td>
<td></td>
</tr>
<tr>
<td>Europe</td>
<td>19</td>
<td>13-28</td>
<td></td>
</tr>
<tr>
<td>Australia/Oceania</td>
<td>2</td>
<td>1-7</td>
<td></td>
</tr>
<tr>
<td>South America</td>
<td>1</td>
<td>0-5</td>
<td></td>
</tr>
<tr>
<td>Multiple Continents</td>
<td>2</td>
<td>1-7</td>
<td></td>
</tr>
<tr>
<td>Unclear</td>
<td>1</td>
<td>0-5</td>
<td></td>
</tr>
<tr>
<td>Objective</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatment</td>
<td>62</td>
<td>52-71</td>
<td></td>
</tr>
<tr>
<td>Prevalence/Incidence</td>
<td>17</td>
<td>11-26</td>
<td></td>
</tr>
<tr>
<td>Economic Evaluation</td>
<td>9</td>
<td>5-16</td>
<td></td>
</tr>
<tr>
<td>Diagnosis</td>
<td>4</td>
<td>2-10</td>
<td></td>
</tr>
<tr>
<td>Other/Unclear</td>
<td>8</td>
<td>4-15</td>
<td></td>
</tr>
<tr>
<td>Study Design</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Meta-Analysis/Network Meta-Analysis</td>
<td>56</td>
<td>46-65</td>
<td></td>
</tr>
<tr>
<td>- Cohort studies</td>
<td>25</td>
<td>18-34</td>
<td></td>
</tr>
<tr>
<td>- Case-Control</td>
<td>1</td>
<td>0-5</td>
<td></td>
</tr>
<tr>
<td>- Cross-Sectional</td>
<td>1</td>
<td>0-5</td>
<td></td>
</tr>
<tr>
<td>- Other</td>
<td>17</td>
<td>11-26</td>
<td></td>
</tr>
<tr>
<td>Type of Prior</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Uninformative</td>
<td>24</td>
<td>17-33</td>
<td></td>
</tr>
<tr>
<td>- Informative</td>
<td>5</td>
<td>2-11</td>
<td></td>
</tr>
<tr>
<td>- Data Driven</td>
<td>4</td>
<td>2-10</td>
<td></td>
</tr>
<tr>
<td>- Not Reported/Unclear</td>
<td>67</td>
<td>57-75</td>
<td></td>
</tr>
<tr>
<td>Software</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- R, SAS, Python, Fortran, S-Plus</td>
<td>42</td>
<td>33-52</td>
<td></td>
</tr>
<tr>
<td>- WinBUGS/JAGS</td>
<td>28</td>
<td>20-37</td>
<td></td>
</tr>
<tr>
<td>- PHASE, BEAST, MrBayes, Genotyping Console</td>
<td>1</td>
<td>0-5</td>
<td></td>
</tr>
<tr>
<td>- Not Reported</td>
<td>3</td>
<td>1-8</td>
<td></td>
</tr>
<tr>
<td>- Other</td>
<td>26</td>
<td>18-35</td>
<td></td>
</tr>
<tr>
<td>Funding Source</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Government</td>
<td>30</td>
<td>22-40</td>
<td></td>
</tr>
<tr>
<td>- Foundation/Association</td>
<td>17</td>
<td>11-26</td>
<td></td>
</tr>
<tr>
<td>- Hospital/University</td>
<td>9</td>
<td>5-16</td>
<td></td>
</tr>
<tr>
<td>- Industry</td>
<td>7</td>
<td>3-14</td>
<td></td>
</tr>
<tr>
<td>- Not Funded</td>
<td>32</td>
<td>24-42</td>
<td></td>
</tr>
<tr>
<td>- Not Reported</td>
<td>21</td>
<td>14-30</td>
<td></td>
</tr>
<tr>
<td>Type of Orthopedic Surgery</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Joint Replacement/Arthroplasty</td>
<td>33</td>
<td>25-43</td>
<td></td>
</tr>
<tr>
<td>- Spine</td>
<td>9</td>
<td>5-16</td>
<td></td>
</tr>
<tr>
<td>- Orthopedic Oncology</td>
<td>7</td>
<td>3-14</td>
<td></td>
</tr>
<tr>
<td>- Shoulder and elbow</td>
<td>5</td>
<td>2-11</td>
<td></td>
</tr>
<tr>
<td>- Sports Medicine/Arthroscopy</td>
<td>2</td>
<td>1-7</td>
<td></td>
</tr>
<tr>
<td>- Trauma</td>
<td>2</td>
<td>1-7</td>
<td></td>
</tr>
<tr>
<td>- Foot and ankle</td>
<td>2</td>
<td>1-7</td>
<td></td>
</tr>
<tr>
<td>- Pediatric Orthopedics</td>
<td>1</td>
<td>0-5</td>
<td></td>
</tr>
<tr>
<td>- Hand and wrist</td>
<td>1</td>
<td>0-5</td>
<td></td>
</tr>
<tr>
<td>- Multiple/general</td>
<td>34</td>
<td>25-44</td>
<td></td>
</tr>
<tr>
<td>- Other</td>
<td>4</td>
<td>2-10</td>
<td></td>
</tr>
</tbody>
</table>

CI – confidence interval.
SAS – statistical analysis system.
WinBUGS – Windows Bayesian Inference Using Gibbs Sampling.
JAGS – Just Another Gibbs Sampler.
PHASE - Program for Reconstructing Haplotypes from Population Data.
BEAST – Bayesian Evolutionary Analysis Sampling Trees.

3.1. ROBUST criteria

We found that studies infrequently reported several items on the ROBUST checklist including, specifying the prior distribution (39%, 95% CI 30–49), justifying the prior distribution (18%, 95% CI 12–27), the sensitivity to different priors (8%, 95% CI 4–15), and the statistical model used (22, 95% CI 15–31). On the other hand, analytical technique (99%, 95% CI 95–100), central tendency (85%, 95% CI 77–91), and variance (81%, 95% CI 72–87) were well reported. Full ROBUST results can be found in Table 2.

Table 2
Number of studies reporting each of the ROBUST criteria.

<table>
<thead>
<tr>
<th>Item</th>
<th>Number reporting n = 100</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prior distribution specified</td>
<td>39</td>
<td>30–49</td>
</tr>
<tr>
<td>Prior distribution justified</td>
<td>18</td>
<td>12–27</td>
</tr>
<tr>
<td>Sensitivity to different priors is examined</td>
<td>8</td>
<td>4–15</td>
</tr>
<tr>
<td>Statistical model explicitly mentioned</td>
<td>22</td>
<td>15–31</td>
</tr>
<tr>
<td>Analytical technique described</td>
<td>99</td>
<td>95–100</td>
</tr>
<tr>
<td>Description of central tendency</td>
<td>85</td>
<td>77–91</td>
</tr>
<tr>
<td>Description of variance as standard deviation or credible interval</td>
<td>81</td>
<td>72–87</td>
</tr>
</tbody>
</table>

CI – confidence interval.

3.2. BayesWatch criteria

Based on the BayesWatch criteria (Table 3), statistical model (22%, 95% CI 15–31), loss function (12%, 95% CI 7–20) prior distribution (37%, 95% CI 28–47) and sensitivity analysis (7%, 95% CI 3–14) were poorly reported. The general methodology items were reported perfectly or nearly perfectly: Interventions (100%, 95% CI 96–100), objective of the study (100%, 95% CI 96–100), design of the study (100%, 95% CI 96–100), and computations (95%, 95% CI 89–98).

3.3. JASP criteria

In the JASP criteria (Table 4), the plot of prior and posterior distribution (0%, 95% CI 0–4), the Bayes factor (2%, 95% CI 1–7), and the results from both estimation and hypothesis testing (4%, 95% CI 2–10) are not reported in almost all the studies. The prior settings (38%, 95% CI 29–48), justification of the prior (18%, 95% CI 12–27) and discussion of the robustness of the results (44%, 95% CI 35–54) are also under reported. The author commonly did not provide enough information on the sequential analysis (2%, 95% CI 1–7), multivariate analyses (5%, 95% CI 2–11), and availability of data files (5%, 95% CI 2–11). Again, general methodological criteria were well reported such as the goal of the analysis (100%, 95% CI 96% - 100) as well as naming the statistical literature referenced (94%, 95% CI 88–97).

4. Discussion

This study demonstrates that many methodological and statistical aspects of Bayesian analysis are consistently underreported in the orthopaedic literature. In particular, details on prior distributions, justification for prior distributions, sensitivity analyses, reporting the Bayes factor, availability of statistical files and data, and the statistical model used in the analysis were consistently poorly reported. Interestingly, we did not find any orthopaedic surgical randomized controlled trials (RCTs) that use Bayesian analyses, although several included studies were secondary analyses of RCT data, indicating underutilization of
Bayesian analyses in orthopaedic clinical trials.

It is possible that one of the reasons so many of the most important aspects of Bayesian analysis are under-reported is due to a lack of understanding of their importance. The Bayesian approach is based on the idea that posterior probability is proportional to prior probability. The prior distribution is at the core of Bayesian analysis. This implies that the results show, most papers don’t specify which prior they are using nor why that prior is being used. Very few papers reported conducting sensitivity analyses to demonstrate the robustness of their choice of prior. Additionally, we found that the statistical model was not reported in many cases. Adequately reporting the statistical model is useful for readers planning to do a similar analysis. It also allows the reader to check that the model was coded correctly. The papers we reviewed consistently left out these key parts of Bayesian analysis. Authors should focus on reporting these key features in the future to improve reporting quality and to enable readers to better understand the analyses.

Sobieraj et al. [9] found that poor reporting quality is common in Bayesian mixed treatment comparisons (MTC). Similar to our paper, Sobieraj found that only 52.9% of MTC studies reported the prior used in the study [9]. Similarly, Brard et al. [10] found that clinical trials using Bayesian survival analysis infrequently reported details on prior distributions. Sung et al. the originators of the ROBUST criteria, conducted a systematic review in 2005 of studies using Bayesian analysis and found that only 3% of studies reported all items on the ROBUST checklist appropriately [6]. They also found that sensitivity to different priors was poorly reported, but they found a higher percentage of studies reporting the prior distribution (78%) than the current study, however their inclusion criteria set limits on the types of priors used so this may have selected for studies that better reported their priors [6].

One strength of our study is that we used three different guidelines to evaluate reporting quality. Although the three reporting checklists identify different items that need to be reported, there was consistency across the checklists in the major aspects of Bayesian analyses; in particular, the importance of giving explicit details on priors and sensitivity analyses were consistently poorly reported across checklists. The wording and instructions of each of the three checklists varies slightly, so it is possible to have a study that meets the criterion on one checklist but not the others. For example, for the prior distribution criterion ROBUST simply asks whether “prior distribution is stated” but BayesWatch has an extensive instruction for that domain and specifically requires that the authors describe whether the prior distribution is informative or non-informative and to state the alternative priors in certain circumstances, so it is possible to meet the criterion on ROBUST but not BayesWatch on this domain. Our review is also strengthened by including various surgical treatments and subspecialties in orthopaedics. However, there were few included studies in several large subspecialties including trauma sports, upper extremity, and pediatrics.

As Bayesian analyses continue to grow in popularity, it becomes increasingly important to develop good reporting practices for better transparency, improved methodological quality, better ability to pool data in meta-analyses and clinical practice guidelines, providing high-quality examples for future investigators, and for replication. We recommend all authors, editors, peer reviewers, statisticians, clinicians, and anyone else that may be involved in the research process for studies using Bayesian analysis to consistently use at least one of the Bayesian reporting guidelines, such as the ROBUST, BayesWatch, or JASP. In doing so, we recommend paying particular attention to key items that are poorly reported including the prior distribution used, the justification, and details on sensitivity analyses.

5. Conclusion

In this paper we identified that there is an opportunity to improve reporting quality and transparency of studies using Bayesian analysis by encouraging adherence to reporting guidelines such as ROBUST, JASP, and BayesWatch. More specifically, there is an opportunity to better report prior distributions, sensitivity analyses, and the statistical models used. We recommend that authors, peer reviewers and journal editors use at least one of the reporting guidelines noted in this review to have a better understanding of reporting of Bayesian analysis.

CRediT author statement

Faris Bdair: Data Curation; Formal Analysis; Investigation; Methodology; Project Administration; Visualization; Writing – original draft
Sophia Mangala: Data Curation; Formal Analysis; Investigation; Methodology; Visualization; Writing – original draft
Imad Kashir: Data Curation; Investigation; Writing – review & editing
Murtaza Shoaib: Data curation; Investigation; Writing - review & editing
Darren Young Shing: Data curation; Investigation; Writing - review & editing
John Price: Data curation; Investigation; Writing – review & editing
Braenne Flood: Data curation; Investigation; Project administration; Writing – review & editing
Samerah Nademi: Data curation; Investigation; Writing – review & editing
Lehane Thabane: Conceptualization; Data curation; Formal Analysis; Investigation; Methodology; Resource; Supervision; Visualization; Writing – review & editing
Kim Madden: Conceptualization; Data curation; Formal Analysis; Investigation; Methodology; Project Administration; Resources; Supervision; Visualization; Writing – original draft.

Conflicts of interest and funding

The authors received no funding for this study. Each author certifies that they have no conflicts of interest in connection with the submitted article.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.
Appendix A. List of included studies


Appendix B. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ctctc.2023.101132.

References