Publishing a Bayesian study – a sometimes overlooked hurdle

Ros Walley, Foteini Strimenopoulou and Andy Grieve

PSI Conference 2018, Amsterdam
Outline

<table>
<thead>
<tr>
<th>Context</th>
</tr>
</thead>
<tbody>
<tr>
<td>Principles for Bayesian reporting</td>
</tr>
<tr>
<td>Potentially contentious topics</td>
</tr>
<tr>
<td>Examples of published studies</td>
</tr>
<tr>
<td>Conclusions</td>
</tr>
</tbody>
</table>
“A drug is a substance which, if injected into a rabbit, produces a paper.”

Otto Loewi

Source of image: https://www.pets4homes.co.uk/pet-advice/vaccinations-in-rabbits.html
Increasing number of Bayesian publications in pharma

Academic publications per year on Bayesian statistics in pharma

Source: SCOPUS
Search: bayesian AND statistic*
"The results are summarised in a bulletin, article, or book, which is essentially a more or less extended memorandum “to who it may concern”.

...<The readers'> purposes, utility functions, prior ideas and other sources of information may differ widely. Uses of the analysis may extend over a considerable period of time after the statisticians work has been completed."
Additional stakeholders at publication stage

New study team members
Internal publications team
Medical writer
Internal governance

Journal editor
Manuscript referees

The ultimate audience....
Journal readership
Principles for Bayesian reporting
Bayesian reporting guidelines

Has a section on “Reporting of Bayesian Analyses of Clinical Trials”

Describes both academic guidance and the regulatory view
Academic guidelines for reporting Bayesian analysis

Bayesian methods in health technology assessment: a review
DJ Spiegelhalter
JP Myles
DR Jones
KR Abrams

Basic Statistical Reporting for Articles Published in Biomedical Journals: The “Statistical Analyses and Methods in the Published Literature” or The SAMPL Guidelines

Thomas A. Lang\textsuperscript{a} and Douglas G. Altman\textsuperscript{b}
\textsuperscript{a} Principal, Tom Lang Communications and Training International
\textsuperscript{b} Director, Centre for Statistics in Medicine, Oxford University

Seven items were identified for inclusion when reporting a Bayesian analysis of a clinical study
Lillian Sung\textsuperscript{a,b,c,d,g}, Jill Hayden\textsuperscript{b,c}, Mark L. Greenberg\textsuperscript{a,c}, Gideon Koren\textsuperscript{a,b,c,d}, Brian M. Feldman\textsuperscript{a,b,c,d,f}, George A. Tomlinson\textsuperscript{f,g,h}
\textsuperscript{a} Division of Hematology/Oncology, Department of Pediatrics, Hospital for Sick Children, Toronto, Ontario, Canada
\textsuperscript{b} Department of Health Policy Management and Evaluation, University of Toronto, Toronto, Ontario, Canada
\textsuperscript{c} Department of Pediatrics, Hospital for Sick Children, Toronto, Ontario, Canada
\textsuperscript{d} Program in Population Health Sciences, Hospital for Sick Children, Toronto, Ontario, Canada
\textsuperscript{e} Institute for Work and Health, Toronto, Ontario, Canada
\textsuperscript{f} Department of Public Health Sciences, University of Toronto, Toronto, Ontario, Canada
\textsuperscript{g} Department of Medicine, University of Toronto, Toronto, Ontario, Canada
\textsuperscript{h} Department of Medicine, University Health Network, Toronto, Ontario, Canada

Accepted 9 August 2004

Draft, 13 September, 2001

Bayesian Standards in Science
Standards for Reporting of Bayesian Analyses in the Scientific Literature

The BaSiS Group

Address correspondence to:
Constantine Gatsonis, gatsonis@stat.tufts.edu, tel: 401 863 9183, or
Steve Goodman: sgoodman@jhu.edu, tel: 410-955-4596.
# Academic Guidelines for Reporting Bayesian Analyses

**ROBUST**

<table>
<thead>
<tr>
<th>Prior Distribution</th>
<th>Specified</th>
</tr>
</thead>
<tbody>
<tr>
<td>Justified</td>
<td></td>
</tr>
<tr>
<td>Sensitivity analysis</td>
<td></td>
</tr>
</tbody>
</table>

**Analysis**

<table>
<thead>
<tr>
<th>Statistical model</th>
<th>Analytical technique</th>
</tr>
</thead>
</table>

**Results**

<table>
<thead>
<tr>
<th>Central tendency</th>
<th>SD or Credible Interval</th>
</tr>
</thead>
</table>

**BAYESWATCH**

<table>
<thead>
<tr>
<th>Introduction</th>
<th>Intervention described</th>
</tr>
</thead>
<tbody>
<tr>
<td>Objectives of study</td>
<td></td>
</tr>
</tbody>
</table>

**Methods**

<table>
<thead>
<tr>
<th>Design of Study</th>
<th>Statistical model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prior / Loss function?</td>
<td>When constructed</td>
</tr>
<tr>
<td>Prior / Loss descriptions</td>
<td></td>
</tr>
<tr>
<td>Use of Software</td>
<td>MCMC, starting values, run-in, length of runs, convergence, diagnostics</td>
</tr>
</tbody>
</table>

**Results**

<table>
<thead>
<tr>
<th>Posterior distribution summarized</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sensitivity analysis if alternative priors used</td>
</tr>
</tbody>
</table>

**Interpretation**

<table>
<thead>
<tr>
<th>Posterior distribution summarized</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sensitivity analysis if alternative priors used</td>
</tr>
</tbody>
</table>

**BASIS**

<table>
<thead>
<tr>
<th>Research Question</th>
</tr>
</thead>
<tbody>
<tr>
<td>Statistical model</td>
</tr>
<tr>
<td>Likelihood, structure, prior &amp; rationale</td>
</tr>
</tbody>
</table>

**Computation**

<table>
<thead>
<tr>
<th>Software</th>
<th>convergence if MCMC, validation, methods for generating posterior summaries</th>
</tr>
</thead>
</table>

**Model checks, sensitivity analysis**

<table>
<thead>
<tr>
<th>Posterior Distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Summaries used: (i) Mean, std, quintiles (ii) posterior shape, (iii) joint posterior for mult comp, (iv) Bayes factors</td>
</tr>
</tbody>
</table>

**Results**

<table>
<thead>
<tr>
<th>Results of model checks and sensitivity analyses</th>
</tr>
</thead>
</table>

**Interpretation of Results**

**Limitation of Analysis**

**SAMPL**

<table>
<thead>
<tr>
<th>Prior Distribution</th>
<th>Specified</th>
</tr>
</thead>
<tbody>
<tr>
<td>Justified</td>
<td></td>
</tr>
<tr>
<td>Sensitivity analysis</td>
<td></td>
</tr>
</tbody>
</table>

**Analysis**

<table>
<thead>
<tr>
<th>Statistical model</th>
<th>Analytical technique</th>
</tr>
</thead>
</table>

**Results**

<table>
<thead>
<tr>
<th>Central tendency</th>
<th>SD or Credible Interval</th>
</tr>
</thead>
</table>
Displaying 6 reporting guidelines found.

Key reporting guidelines, shaded green, are displayed first. Show the most recently added records first.

1. STARD-BLCM: Standards for the Reporting of Diagnostic accuracy studies that use Bayesian Latent Class Models

2. Latent Class Analysis: An example for reporting results

3. An introduction to using Bayesian linear regression with clinical data

4. Seven items were identified for inclusion when reporting a Bayesian analysis of a clinical study

5. Bayesian methods in health technology assessment: a review

6. Point and interval estimates of effect sizes for the case-controls design in neuropsychology: rationals, methods, implementations, and proposed reporting standards
More recent publications on guidelines for publishing

Minimum reporting guidelines:

- Complete description of likelihood and all priors
- Details of convergence
- Software details
- Provide data and scripts where possible
- Be explicit how posterior is summarised
- Report no of draws from the posterior as well as the effective sample size
More recent publications on guidelines for publishing

Specific context of network meta-analysis for safety but some good general points, including:

- Rationale for choice of prior; justification of exchangeability assumption; when prior was specified
- Stats model: rate of missingness; imputation of missing covariates; structure of levels of model....
- Impact of prior on posterior
- Possible limitations of the analysis
Decision/success criteria

- Clinical studies typically have decision criteria associated with them
  - May not be explicitly stated. E.g. safety objectives, criteria related to $p<0.05$
- Bayesian guidelines do not focus on decision criteria
- Decision criteria for internal benchmarking may not be “set in stone”
  - Leads to a reluctance to document and publicise

However,

- Studies need a sample size justification
  - So typically at least one decision criterion is at least implied by the protocol
- If decision criteria are included in the protocol or SAP, then
  - Easy to demonstrate they were “prespecified”
  - Thus easier to convince sceptical reviewers
Potentially contentious areas
Comments from stakeholders at publication stage

The Bayesian method hasn’t worked (the estimated treatment diff doesn’t match the summary stats)

Can we include p-values?

Statistical analysis should be performed paired and unpaired fashion for all data

NB. Not all these issues are specifically Bayesian
The discussion of the prior distribution should also show the treatment group and how the prior impacted its shift.

We are extremely limited by the number of additional words that we can add to the main body of the manuscript.

There is nothing wrong with Bayesian approaches to analyses, but the choice of the parameters for the priors is unsupported.

Statistical analysis should be performed paired and unpaired fashion for all data.

Can we include p-values?

The Bayesian method hasn’t worked (the estimated treatment difference doesn’t match the summary stats).

NB. Not all these issues are specifically Bayesian.
Not all decision-criteria were described in protocol

Statistical analysis should be performed paired and unpaired fashion for all data.
Can we include p-values?
The Bayesian method hasn’t worked (the estimated treatment diff doesn’t match the summary stats)

Standard of care threshold is now out of date and we don’t want to mention it

We are extremely limited by the number of additional words that we can add to the main body of the manuscript

There is nothing wrong with Bayesian approaches to analyses, but the choice of the parameters for the priors is unsupported

Need to justify a non-zero threshold based on standard of care (in-house assessment of a competitor)

Statistical analysis should be performed paired and unpaired fashion for all data.

The Bayesian method hasn’t worked (the estimated treatment diff doesn’t match the summary stats)

We are extremely limited by the number of additional words that we can add to the main body of the manuscript

There is nothing wrong with Bayesian approaches to analyses, but the choice of the parameters for the priors is unsupported

Need to justify a non-zero threshold based on standard of care (in-house assessment of a competitor)

NB. Not all these issues are specifically Bayesian
More complex analysis can be viewed with suspicion

**Dynamically downweighting priors e.g. mixture priors**
- “Are you changing the prior after the study has started?”

**Extending the likelihood to downweight outliers e.g using a t-distribution or mixture likelihood**
- “Which outliers have been excluded?”

**Assessment of the design is more complex than in the frequentist setting. E.g.**
- type 1 error/bias when there are informative priors
- performance of dynamically downweighting priors
- Assessment of prior data conflict
- “too much detail” or “what are you hiding?”
Examples: excerpts from published papers
Summary

Methods: The primary efficacy endpoint was the percentage of patients with a 20% response according to the Assessment of SpondyloArthritis international Society criteria for improvement (ASAS20) at week 6 (Bayesian analysis).

Findings: At week 6, ASAS20 response estimates were 59% on secukinumab versus 24% on placebo (99.8% probability that secukinumab is superior to placebo)
Abstract: Using the predefined primary assessment of efficacy (Bayesian analysis with informative prior), we observed ....

Statistical analysis:

• A two-part decision criteria for efficacy and futility was used at the end of this study which is described in the Supplemental Material

• More details of the statistical design and analysis including the outlier robust approach can be found in ....
Statistical methods:

The analyses of ACRn and ACR20 at week 8 followed the Bayesian paradigm to improve the operating characteristics of the study design. The Bayesian methodology was used for the analysis of ACRn and ACR20 at week 8 assuming informative priors only on the placebo response (vague prior distributions were assumed for all other parameters). For both endpoints, the prior placebo response, in terms of ACR20 response rate, was approximately 25% and the effective sample size was 32. Normal likelihood model and logistic model were assumed for ACRn and ACR20, respectively.
Statistical methods:

Two study efficacy criteria were predefined based on ACRn at week 8. For declaring PoC, we required a high (≥97.5%) posterior probability that bimekizumab is superior to placebo. To give further confidence about the effect size, the posterior probability that bimekizumab improvement over placebo exceeds a clinically relevant effect (ie, approximately 25% difference from placebo in ACR20 terms) was required to be ≥70%.
Conclusions

There are additional stakeholders at publication stage
- They don’t have the benefit of understanding/inputting to the design in advance of the study
- May be expecting a frequentist discussion
- They may be suspicious/inexperienced with Bayesian methods
- Include a respected stats expert for the contentious points

Consider the potential publication(s) at the design stage
- Prespecify and document as much as possible
- Consider alternative priors/demonstrate robustness of conclusions to these
- Can’t easily demonstrate pre-specification if it's not in the protocol or SAP
- Clear wording in protocol and SAP
- Consider publishing design in stats journal (more sympathetic to technical details)

In the publication
- Follow available guidelines for publication
- Be clear what’s been pre-specified
- Be clear what's the primary analysis and what's not
- Warn other authors that glossing over details may lead to suspicion
- If authors won't agree, at least put the wording in the supplemental documents
References (1)


Glatt S, Baeten D, Baker T, et al. Dual IL-17A and IL-17F neutralisation by bimekizumab in psoriatic arthritis: evidence from preclinical experiments and a randomised placebo-controlled clinical trial that IL-17F contributes to human chronic tissue inflammation. Annals of the Rheumatic Diseases 2018;77:523-532


References (2)


### Some differences between reporting frequentist and Bayesian analysis

<table>
<thead>
<tr>
<th></th>
<th>Frequentist</th>
<th>Bayesian</th>
</tr>
</thead>
<tbody>
<tr>
<td>Readership</td>
<td>Familiarity with methods</td>
<td>May be less familiar and even suspicious</td>
</tr>
<tr>
<td>Success criterion</td>
<td>Assumed to be 5% statistical significance for primary comparison (based on a 2-sided test)</td>
<td>No such assumption</td>
</tr>
<tr>
<td>Prior belief</td>
<td>Described in intro and discussion</td>
<td>Specified mathematically</td>
</tr>
<tr>
<td>Approach</td>
<td>Typically use analytical results e.g. Transform and remove outliers to assume normally distributed data</td>
<td>Can easily use different distributions for the model. No particular driver for simpler models.</td>
</tr>
<tr>
<td>Estimand</td>
<td>Summary measures relates to analytical approach e.g. odds ratio in logistic regression</td>
<td>No restriction</td>
</tr>
</tbody>
</table>
Bayesian reporting guidelines

Describes FDA reviewers perspective of reviewing Bayesian medical device trials
N.B. Refers to a regulatory review rather than a peer review for a journal