Developing critical thinking about reporting of Bayesian analyses

Eleanor M. Pullenayegum
Qing Guo
McMaster University
St Joseph's Healthcare Hamilton

Robert B. Hopkins
St Joseph's Healthcare Hamilton

Journal of Statistics Education Volume 20, Number 1 (2012),
www.amstat.org/publications/jse/v20n1/pullenayegum.pdf

Key Words: Active learning; Writing; Graduate education; Biostatistics; Bayesian methods

Abstract

Graduate students in the health sciences who hope to become independent researchers must be able to write up their results at a standard suitable for submission to peer-reviewed journals. Bayesian analyses are still rare in the medical literature, and students are often unclear on what should be included in a manuscript. Whilst there are published guidelines on reporting of Bayesian analyses, students should also be encouraged to think about *why* some items need to be reported whereas others do not. We describe a classroom activity in which students develop their own reporting guideline. The guideline that the students produce is not intended to replace existing guidelines, rather we have found that the process of developing the guideline is helpful in encouraging students to think through the "why?" as well as the "what?" of reporting.

1. Introduction

Graduate students preparing for careers in academic research must write up their results at a standard suitable for submission to peer-reviewed journals; there is little use in students being able to do a statistical analysis if they cannot disseminate their findings. Writing the results requires students to know both how and what to write. Students benefit from guidance on both these issues, but the focus of this paper is on the "what" rather than the "how" (see Becker,

Richards & Ebooks Corporation 2007, Boice 1990 for advice on the "how"). We focus here on Bayesian analyses, which we teach in a graduate course designed for students specialising in public health, health economics, clinical epidemiology, or biostatistics.

Students often ask how to report the results of statistical analyses, for example whether they need to include checks for normality. Since standard statistical analyses are ubiquitous in the medical literature, students can get a good sense of what is usually reported. Since Bayesian analyses are far less widely used in medicine, it is difficult for students to gauge what is typical. Thus, teaching students how to report Bayesian analyses in the medical literature is an important component of a course in Bayesian biostatistics.

When teaching reporting, we need to avoid training students to follow a set of rules unthinkingly. Rather we need to focus on helping students to think critically about what is needed. That is, we need to teach not just the "what", but also the "why". There are a number of published guidelines detailing what should be reported when writing papers for the medical literature. These include specific guidelines on Bayesian analyses (Spiegelhalter, Abrams & Myles 2004, Sung, Hayden, Greenberg, Koren, Feldman & Tomlinson 2005, The BaSiS Group 2001). It can be tempting to teach reporting by pointing students to a published guideline and asking them to use it on a sample manuscript. The difficulty with this approach is that it asks students to apply a set of rules, but does not ask them to evaluate why those rules are reasonable.

This raises the question of how to teach reporting in such a way that students know what to report, and also understand why some items are important but others are not. In this paper, we describe a classroom activity that we have found helpful in conveying both the "what" and the "why". Students are asked to develop their own reporting guideline, using a process similar to that used in developing the guidelines in the medical literature (e.g. Begg et al. 1996, Moher, Cook, Eastwood, Olkin, Rennie & Stroup 1999). In a modified Delphi method, participants provide anonymous suggestions on items to report, and these form an initial list. Each item on this initial list is discussed, and anonymous feedback is used to establish the final list. Anonymity in both stages of feedback is intended to avoid dominant individuals having undue influence.

The purpose of this paper is to describe the classroom activity. The point of the exercise was not the final guideline that students produced, but rather to have students go through the process of creating a guideline as a class, thus being forced to justify opinions as to what is or is not important to report in a manuscript.

2. An overview of the activity

Before the exercise began, the class agreed to treat every suggestion with respect. Initially, students were asked to write a list of everything they thought should go into a report of a Bayesian study. Our class was small (six students), allowing us to collate the lists using pen and paper before sharing using a laptop and data projector. Students took a heading (e.g. methods, priors, results, other) and noted all the suggestions falling under that heading. Anonymity in producing this initial list was effective in involving everyone. Having the instructor participate anonymously can force certain items into the discussion. In our case, the instructor anonymously

put initial values, priors for monitoring, and convergence on the list. She also included some items that are not crucial, for example history plots from Gibbs sampler simulations.

Students discussed each item, making the case for why it was or was not important. For each item, we considered how its inclusion might help to understand or interpret the results, or to verify whether the analysis was done correctly. Since the course was on Bayesian methods, students were told that the guideline they were developing was intended as a supplement to existing guidelines for specific study designs (e.g. Schulz, Altman, Moher, and CONSORT Group 2010 for clinical trials or, von Elm, Altman, Egger, Pocock, Gotzsche & Vandenbroucke 2008 for observational studies). Thus we deleted items such as sample size or study objectives, since these are covered in general guidelines. This focused discussion on Bayesian methods rather than on research methods in general. Overlapping items were combined, as when items are very similar it is difficult to rate the importance of each item independently (for example, if "graphs" and "posterior density plots" are both on the initial list, "posterior density plots" may be judged not important if "graphs" is already included). This formed the initial list.

In the final step, the list of candidate items was refined to include only those considered necessary. Using a three-point scale ("not important at all", "useful to report if space allows", and "don't publish the manuscript without this"), students anonymously rated each item. These ratings were relayed to the moderator using pen and paper. We found the mode and range of ratings for each item to be the quickest way to summarize; students can each take an item to speed up the process. When results are collected electronically, it is more informative to report the percentages in each category. The objective of the Delphi process is to achieve consensus, so if at this point there is disparity in the responses, it may be helpful to discuss again to resolve any misunderstandings in wording or overlap in items, and then repeat the anonymous ratings. We did not find this necessary in our classes. For our final checklist, we included only those items for which "don't publish the manuscript without this" was the most common response.

This exercise was done with classes in both 2009 and 2010. After the 2010 class students were asked to write a reporting guideline, presenting the class checklist and explaining the reasons for including each item. The class was told beforehand that this assignment would be given, hoping that doing so would help each student to engage in the discussion, as every student would become responsible for justifying the final product.

3. The class checklist

We now describe the checklists our class proposed, and the reasons for their choices. We begin with the initial list of items, clarify their meaning where necessary, summarize the class discussion of the importance of each item, and then present the final checklist. For simplicity, we describe the process for 2009 only; the process for 2010 was similar.

The purpose of the exercise was not to add to the existing literature on how to report Bayesian methods (for published guidelines, see Spiegelhalter, Abrams & Myles 2004, Sung et al. 2005, The BaSiS Group 2001). Rather, the value of the exercise is in students learning why some items are important and others not. The measure of success is not whether or not the guideline is

reproducible, but rather whether students were able to identify which items should be reported and give a justification for their choices.

3.1 Initial list

Our initial list of items for consideration consisted of the following:

Priors

(a) **How prior was constructed from prior information**, that is, how a distribution and its parameters are fitted to the prior data. For example, if a Normal distribution was chosen we may choose to match on the mean and variance. Similarly, if we have a number of prior distributions elicited from experts, we may construct an overall prior by arithmetic pooling (Spiegelhalter, Abrams & Myles 2004).

(b) **When the prior was constructed**. For example, this could be before data collection, during data collection but before analysis, or after data collection but before analysis.

(c) **Prior distributions used**, that is, the parametric distribution used to describe the prior, e.g. Normal, Beta, or binomial.

(d) **Parameters for the prior distributions**. For example, if the distribution were Normal, we would need to indicate the mean and precision to specify the prior distribution completely.

(e) **Sources of prior information.** This could include systematic reviews, previous studies, or expert opinion.

(f) **Informative priors,** i.e. priors that incorporate existing knowledge.

(g) **Non-informative priors** or vague priors, designed to let the data dominate.

(h) **Prior sensitivity analysis**. For example using a distribution from a different parametric family or, more commonly, varying the parameters of the distribution.

(i) **Non informative and informative priors with sensitivity analysis**. This was not initially listed, but following the discussion students chose to introduce this and remove the items "informative priors", "non-informative priors" and "sensitivity analysis" in order to avoid overlap between the items.

(j) **Whether sensitivity analyses are comprehensive.**

Posterior derivation

(a) **Whether the posterior distribution was derived by simulation or by analytical methods.** This was not initially on the list, but was added during the discussion since other items such as iterations, initial values and convergence are not relevant if the posterior is derived analytically.

(b) **Number of simulations.** For analyses in which the posterior distribution is derived through MCMC simulations, the number of simulations performed.

(c) **Initial values used.** The starting point for a MCMC simulation.

(d) **Methods of monitoring convergence and their adequacy**. MCMC simulations are set up so that their stationary distribution is the posterior distribution of interest. Hence, when the chain has converged, we will be sampling from the posterior distribution.

(e) **Number of iterations burned in.** MCMC simulations drawn before the chain has converged do not represent the posterior distribution, and hence should be discarded, or "burned in".

(f) **Software used, if any** (e.g. WinBUGS, R, SAS).

Model

(a) **Likelihood functions.** The likelihood function for the data that will be combined with the prior in order to derive the posterior distribution.
(b) **Model code,** e.g. WinBUGS model statement.

Other methodological considerations

(a) **Rate of missingness and whether missing covariates were imputed.** Missing outcomes are imputed automatically in WinBUGS, whereas imputation of covariates requires additional code.
(b) **Plans for monitoring** This was intended specifically for the case of Bayesian trial design in which interim monitoring for efficacy or futility is planned, with a view to stopping the trial early if indicated.
(c) **Sample size calculation.** Justification for the number of individuals enrolled in the study.

Presentation of Results

(a) **Graphs: Triplot.** The triplot is a plot of the prior distribution, the likelihood function, and the posterior distribution. Often this is included to show the relative contribution of prior and data to the posterior.
(b) **Graphs: Posterior density function** (or histogram).
(c) **Results – an estimate of central tendency (e.g. mean or median) and spread (e.g. 95% credible interval or standard error) for parameters of interest.**
(d) **Results of convergence tests.**
(e) **A statement about whether any evidence for non-convergence was found.** This was not initially on the list, but students chose to add it during their discussion.

3.2 Summary of discussion

Priors

Knowing how a distribution was fitted to the prior evidence is valuable, but not as vital as knowing what the prior actually was. When the prior was constructed (i.e. before, during, or after data collection) is relevant as there is potential for bias when this is done after data collection begins, particularly if the evidence is based on expert opinion. However, in many cases it may be reasonable to revise a prior over the course of the study if other external results become available. Thus, it is the source of external evidence, rather than the timing, that is critical. Sensitivity analysis is important since readers often differ in their prior beliefs. In our class, we approached sensitivity analysis through enthusiastic, non-informative and sceptical priors (Spiegelhalter, Myles, Jones & Abrams 2000). In this approach, before the study is begun three priors are formed: a prior describing the beliefs of a group of investigators who believe that the effect of interest is present (the enthusiastic prior), a prior describing the beliefs of a group of reasonable sceptics about the effect of interest (the sceptical prior), and the prior to be used for analysis (which would usually lie between the sceptical and enthusiastic priors). A non-informative prior may also be used: it is often helpful to know what evidence the data contributes on its own. Whilst other, more mathematical, approaches to sensitivity analysis are possible, this approach conveys the rationale of sensitivity analyses to non-statisticians. Sensitivity analyses should be comprehensive, as only if a reader believes a prior is he likely to believe the posterior; however, whether they are comprehensive is a judgement best left to the reader.

Posterior derivation

Arguments in favour of reporting the number of simulations, the number of iterations burned in, and the initial values were that these would be needed if a reader wished to replicate the analysis exactly, and that if the total number of iterations required were large and another analyst wished to replicate the results, it would be courteous to warn them of the computation time. However, since any analyst wishing to replicate the results would need the data, the authors could warn of this upon being contacted for the data. Moreover, the number of iterations, the burn-in, and initial values do not help to interpret the results provided the reader believes that the chain has converged with an appropriate burn-in phase.

It was agreed that monitoring convergence was important. However, since results for a non-converged chain should not be published, any reports of convergence diagnostics would not show evidence of non-convergence. This led to a suggestion to include in the results "A statement about whether any evidence for non-convergence was found". This is not ideal, as it is possible that no evidence for non-convergence was found simply because no evidence was sought. However, in the final rating students chose not to include the item "methods for monitoring convergence", so evidently they felt that in making this statement it was implicit that some diagnostics had been performed.

Model

Model code is useful for readers planning to do a similar analysis. It also allows the reader to check that the model was coded correctly. However, since code occupies a lot of space, it could be included in a web appendix.

Posterior description

Although "graphs" was initially listed without qualification, students subsequently felt this was too vague. History and autocorrelation plots are unlikely to be helpful relative to the space they occupy, leaving triplots and density plots of the posterior distribution as candidate graphs. Triplots illustrate the influence of the prior; however, this can also be illustrated by comparison to the results with a non-informative prior. The posterior density plot is valuable because the product of a Bayesian analysis is a distribution, not a point estimate. However, when the posterior distribution of interest is Normal, giving an estimate of location and an estimate of spread together with the distributional family would adequately summarize the distribution.

Other methodological considerations

Sample size was raised since sample size considerations for Bayesian methods are different from frequentist methods. It was noted that sample size is included in many study design-specific guidelines already. Monitoring should be included in any report of a randomized controlled trial, and so the importance of including this in a Bayesian checklist is questionable.

After this discussion students rated each item, and only those items for which the most frequent response was "don't publish the manuscript without this" were retained. Our final checklist is given in Table 1 under the 2009 heading.

Table 1: The published ROBUST checklist, and the class checklists from 2009 & 2010

ROBUST	2009 checklist	% reporting (n/N)	2010 checklist	% reporting (n/N)
Priors				
Specified	Distribution(s) used	83% (5/6)	Existing knowledge/results from previous studies/experts'	75% (3/4)
Justified	Values of distributional parameters	83% (5/6)		
Sensitivity analyses	Source for prior information	100% (6/6)	How the prior distribution is constructed (e.g. by matching the means/percentiles)	75% (3/4)
	Non informative and informative priors with sensitivity analysis	100% (6/6)	Reasons for using a vague/non-informative prior, if applicable.	100% (4/4)
Analysis				
Statistical model	Likelihood function, types of analysis	83% (5/6)		
Analytic technique	Whether the posterior was derived by simulation or analytical methods	100% (6/6)		
	Software used, if any	100% (6/6)		
	Rate of missingness and imputation of any missing covariates	50% (3/6)		
Results				
Central tendency	Central tendency + spread (95% CrI/se)	100% (6/6)	Central tendency + 95% CrI	100% (4/4)
Spread/precision: std deviation/credibility interval	A statement about whether any evidence for non-convergence was found	100% (6/6)	Whether problems with convergence were found (if yes, specify)	100% (4/4)
			Comparison of the results to the prior information	75% (3/4)
			Limitations, e.g. missing data, uncertainty in convergence.	50% (2/4)
			Generalizability.	75% (3/4)

After the exercise, students were shown the most recent published guideline (ROBUST; Sung et al. 2005) for comparison. The 2009 checklist contains all the elements of the ROBUST guideline, plus software, missingness, and non-convergence. The BaSiS and Bayeswatch guidelines both suggest reporting which checks for non-convergence were done, rather than simply whether evidence of non-convergence was found. They also include software. None of the other guidelines mentions missingness. Our checklist requests that the source of prior information be specified; others (Sung et al. 2005, The BaSiS Group 2001) simply request that the prior be justified. Whilst all the guidelines request sensitivity analyses, ours is the only one that requests non-informative priors. Bayeswatch (Spiegelhalter, Abrams & Myles 2004) requests the initial values and the number and length of runs.

In summary, our checklist was similar to previous lists, but different enough to indicate that students were willing to think for themselves.

4. Measures of Learning

We have measures of learning one month after the exercise for 2009 and 2010. For 2010, we also have indicators of learning measured during the class, and within a week of the class.

4.1 In-class measures

Learning during the class can be measured by comparing items suggested on the initial list to those items with a consensus rating as "essential" on the final list. In 2010 the set of initial items was kept separate for each student, and we can thus assess whether some items initially viewed as important by most students were subsequently dropped, or whether some items left off most students' initial lists were subsequently rated as important. Due to the anonymous nature of the process, we cannot assess changes at the individual level.

The lists of items initially suggested by each student are included in the Appendix, and changes in opinions can be assessed by comparing to the final list in Table 1. For example, two of the students (and the instructor) requested details of how convergence was assessed on the initial list. After discussion and rating, there was agreement that this was helpful to put in if space allowed, but not essential. Similarly, only one student mentioned reporting whether there was any disparity between prior information and the data, however this was included in the final set. We acknowledge that this is a limited measure of learning, as the class was small (there were 4 students enrolled, one of whom was absent the day of this particular class), and we cannot assess changes at the individual student level.

4.2 One-week post-class

Learning immediately after the class was measured by an assignment in which students were asked to present their checklist and justify why each item was important (this was done in 2010 only). Students provided good reasons for each of the items in the prior section. They all explained that reporting how the prior was constructed was important in ensuring a reproducible analysis. Each student also identified a reason for comparing the results to the prior information; the two primary reasons are firstly to describe the new knowledge added and characterize shifts

in evidence, and secondly to identify any major prior-data conflicts that may call the validity of the posterior into question.

Justification for items in the analysis and results sections was less adequate. For example, only one of three students adequately explained that when MCMC methods are used for analysis, the resulting samples describe the posterior only if the chain has converged. All students identified a good reason for reporting on generalisability, but only one commented on the specifically Bayesian consideration that the generalisability of the results depends on to what degree readers agree with the prior(s) used.

Although the numbers are small, this suggests that all students had a good grasp of why each element of prior construction and comparison to the posterior distribution was important. Not all students provided full justifications of the remaining four items. In cases where a full justification was not provided, there was usually a partial justification. At times it was difficult to tell whether an inadequate justification was due to poor understanding or due to language issues (only one student was a native English speaker).

4.3 One month post-class

The final measure of learning was students' written project reports. In 2009 students submitted a written report, whilst in 2010 they submitted a written report which was critiqued by a partner and handed back for corrections before final submission. Table 1 outlines the percentage of students reporting each item from their own checklist. In 2010, peer reviewers identified all omissions from their checklist, and these were corrected on the final submission. This shows that one month after the exercise, students retained a good grasp of which items to report.

4.4 Did attitudes change?

Two of us (RBH and QG) participated in the first implementation of the exercise as students. As the exercise progressed, we found ourselves thinking as readers, reviewers or editors rather than as the statistician or author. We also experienced the process of developing a guideline, for example how items were chosen at the end. Actually doing it was a deeper learning experience than just reading it. It was helpful that this was done over a short period of time in class. If this was given as homework, then the thinking would not happen because students would search online for lists.

4.5 Timeline on activity

We used two hours to introduce and complete the exercise. In an attempt to streamline the process, in 2010 students were asked to submit an initial list of items through an anonymous survey on the course's e-learning page. It was hoped that this would allow us to generate the initial list before the class, and also to set up a second survey which students could use during class to rate the importance of each item. However a major server failure left the university's e-learning system unavailable the day before the class, rendering this impossible. As technical glitches are not uncommon, we suggest allowing two hours for the activity.

5. Discussion

We have described a class activity intended to help students identify important items to report in a Bayesian analysis, with an emphasis not just on what to report, but on understanding why some items are more important than others.

Several limitations should be noted. Since our classes were small (class sizes of 6 and 4 in 2009 and 2010, respectively), data on learning is limited. Moreover, it is easier to achieve consensus in a small group. Larger classes may lead to a richer discussion, however as class sizes grow it becomes harder for everyone to participate. This may result in less consensus and consequently less ownership of the finished product. Larger classes may need to split into groups, perhaps moderated by a teaching assistant, with each group producing its own list.

The role of the instructor in the process deserves some consideration. The entire process could be left in the hands of the students, or the instructor can be participating on an equal footing with students. From our perspective, it is helpful to have the instructor facilitating, since students will not be familiar with the process. We have also found that with very small and friendly classes, students can agree with each other before considering all aspects, and thus the instructor may at times need to play "devil's advocate". Further, with small classes it is helpful for the instructor to add items (anonymously) to the initial list, for two reasons. Firstly, it is possible that all the students initially leave an important item off the list. When this happens, the process is such that the item would not be discussed at all, and would not be on the final list. Secondly, keen students may have read the published guidelines before class, and so only include items from the published guidelines. When this happens, students do not consider why other items are not important. The instructor can facilitate the learning process by ensuring that all the essential items make it onto the initial list, and also that some less important items are included.

We have used this activity with health sciences graduate students who typically specialise in epidemiology, biostatistics, health economics or public health. These students are familiar with the concept of reporting guidelines and part of the appeal of the exercise is experiencing the process of creating such a guideline. Such guidelines may not exist in other disciplines, or may be created using different methods. The relevance and appeal of the activity is dependent on students' backgrounds.

Finally, we do not intend that the guidelines our classes produced replace the existing guidelines on reporting Bayesian biostatistics. We acknowledge that our methods will not necessarily produce the same guideline each time. At the end of the exercise we showed students the existing guideline, describing it as the current reporting standard. Thus, reproducibility of the guidelines emerging from the process is not critical. The value of the activity is in the learning process rather than the guideline itself.

In terms of achieving our goals of conveying both the "what" and the "why" of reporting, we had some success at conveying the "why" and good success at conveying the "what". Student assignments immediately after the exercise showed good understanding of the "why" for half the items included on their checklist, but incomplete understanding for other items. Critical thinking could be further enhanced by identifying items that were inadequately justified and asking

students to submit a revision, since at times it appeared that students had spent more time explaining what each item meant and may simply have forgotten to include a justification. In their written end-of-term projects, submitted one month after the exercise, students reported the important items well, showing good adherence to the current standard for reporting Bayesian analyses in biostatistics.

We have found this activity to be an engaging way to teach reporting of Bayesian methods, and hope it will be helpful to other instructors.

Appendix
Initial lists suggested by students and instructor in 2010

Student 1

- What the existing knowledge/results from previous studies/experts' opinions are.
- Specifically how the prior distribution is constructed (i.e. by matching the means/percentile, or other methods of elicitation).
- If an (sic) vague/non-informative prior were used, what the reasons are.
- Descriptions of the data, i.e. variables, type of variables, etc.
- What software (and its version) is used for computing the posterior distribution and what the codes are.
- The details of how the convergence is assessed (i.e.,diagnostic tests (sic), number of iterations, etc.)
- The interpretations of the results and the comparison with the prior information.
- Any limitation of the analysis, i.e. missing data, uncertainty in the convergence, etc.

Student 2

- Priors for mean/proportions and where they come from.
- Priors for precision and where they come from.
- Stopping rules, incorporating enthusiastic and sceptical priors.
- Results to be reported as means/proportions with 95% credence intervals.

Student 3

- Hypothesis of interest and primary/secondary outcomes.
- Target population.
- Sampling method/procedure and size of the sample.
- Response rate and flow diagram for participnats (sic) recruited/refused.
- Methods for statistical analysis:
    - Specification and justification for the distribution of outcome/response.
    - Specification and justification for the prior distribution and hyper-parameters.
    - Posterior distribution (if possible analytically).
- Result:
    - Number of itteration (sic) and number of simulation run.
    - Convergence diagnostics and correlations.
    - Posterior distribution characterized by Mean/Median and credible interval.
- Conclusion:
    - Main findings.
    - Generalizability.
    - Limitations.

Instructor

- Software used, if any
- Initial values
- Tests of convergence used
- Results of convergence tests
- Priors used
- Information on which the priors were based
- How the prior was constructed from prior information
- Number of iterations burned in
- Total number of iterations
- Number of chains run
- WinBUGS code, if any
- Descriptive stats on summary distributions
- MCMC error
- Likelihood function
- Posterior correlation between parameters
- Triplots

References

Becker, H.S., Richards, P. & Ebooks Corporation (2007), *Writing for Social Scientists: How to Start and Finish Your Thesis, Book, or Article*, The University of Chicago Press, Chicago.

Begg, C., Cho, M., Eastwood, S., Horton, R., Moher, D., Olkin, I., Pitkin, R., Rennie, D., Schulz, K.F., Simel, D. & Stroup, D.F. (1996), "Improving the quality of reporting of randomized controlled trials. The CONSORT statement", *JAMA : the Journal of the American Medical Association,* vol. 276, no. 8, pp. 637-639.

Boice, R. (1990), *Professors as writers : A self-help guide to productive writing*, New Forums Press, Stillwater, Okla., U.S.A.

Moher, D., Cook, D.J., Eastwood, S., Olkin, I., Rennie, D. & Stroup, D.F. (1999), "Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement. Quality of Reporting of Meta-analyses", *Lancet,* vol. 354, no. 9193, pp. 1896-1900.

Schulz, K.F., Altman, D.G., Moher, D. & CONSORT Group (2010), "CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials", *British Medical Journal (Clinical research ed.),* vol. 340, pp. c332.

Spiegelhalter, D.J., Abrams, K.R. & Myles, J.P. (2004), *Bayesian approaches to clinical trials and health care evaluation,* Wiley, Chichester ; Hoboken, NJ.

Spiegelhalter, D.J., Myles, J.P., Jones, D.R. & Abrams, K.R. (2000), "Bayesian methods in health technology assessment: a review", *Health technology assessment (Winchester, England)*, vol. 4, no. 38, pp. 1-130.

Sung, L., Hayden, J., Greenberg, M.L., Koren, G., Feldman, B.M. & Tomlinson, G.A. (2005), "Seven items were identified for inclusion when reporting a Bayesian analysis of a clinical study", *Journal of clinical epidemiology*, vol. 58, no. 3, pp. 261-268.

The BaSiS Group (2001), September 13 2001-last update, *Bayesian standards in science (BaSiS)*. Available: http://lib.stat.cmu.edu/bayesworkshop/2001/BaSis.html.

von Elm, E., Altman, D.G., Egger, M., Pocock, S.J., Gotzsche, P.C., Vandenbroucke, J.P. & STROBE Initiative (2008), "The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies", *Journal of clinical epidemiology*, vol. 61, no. 4, pp. 344-349.

Eleanor M. Pullenayegum
Dept of Clinical Epidmiology & Biostatistics, McMaster University
Biostatistics Unit, St Joseph's Healthcare Hamilton
50 Charlton Ave E
Hamilton, ON, L8N 4A6
Canada
mailto:pullena@mcmaster.ca

Qing Guo
Dept of Clinical Epidmiology & Biostatistics, McMaster University
Biostatistics Unit, St Joseph's Healthcare Hamilton
50 Charlton Ave E
Hamilton, ON, L8N 4A6
Canada
mailto:guoq@mcmaster.ca

Robert B. Hopkins
PATH Research Institute
Suite 2000, 25 Main Street W
Hamilton, ON, L8P 1H1
Canada
mailto:hopkinr@mcmaster.ca
