Project Name: Use of Bayesian Techniques in Randomized Clinical Trials: A CMS Case Study
Project ID: STAB0508

Disposition of Comments

Table 2: Public Review Comments

Reviewer Name[1]	Reviewer Affiliation[2]	Section[3]	Reviewer Comments	Author Response[4]
Jose Ma. J. Alvir, DrPH	Pfizer, Inc	Methodology	The TA should further clarify how Bayesian models are applied differently as compared to classical models. The TA provides a thoughtful analysis of the benefits and drawbacks of using a Bayesian analytic approach versus a classical (frequentist) statistical approach to data analysis. The TA concludes that "direct comparisons of meta-analysis between frequentist and Bayesian approaches do not always yield consistent results."[AHRQ. "Use of Bayesian Techniques in Randomized Clinical Trials: A CMS Case Study." Accessed at http://www.ahrq.gov/clinic/ta/bayesian.pdf on July 6, 2009. Page 52.] We recommend that AHRQ provide greater clarification and discussion why results from classical and Bayesian approaches can differ and what the implications would be for policy decisions. If CMS is to adopt Bayesian analysis in its decision-making, the agency will need guidance on how to consider and incorporate results that are conflicting, in addition to assessing the appropriateness of model selection in each case. We also recommend the TA clarify the terms under which a fixed-effect versus random-effect model should be used when incorporating prior data in Bayesian models. Currently, interpretation of statements in the TA could imply that fixed-effect models are preferable because they give weight to larger studies. However, this implication does not necessarily account for potential heterogeneity, which requires downweighting of the larger studies. To help clarify these issues, the TA should better address the issue of heterogeneity related to both classical and Bayesian model use.	We now include in the Tutorial a new discussion of fixed- and random-effects models (under frequentist and Bayesian approaches), heterogeneity, and interaction (page 12).

Jose Ma. J. Alvir, DrPH	Pfizer, Inc	Implemen-tation	Appropriate infrastructure and workforce is needed to implement use of Bayesian statistics in CMS decision-making. We agree with the TA's discussion of some of the potential disadvantages of Bayesian method implementation, such as the lack of statistical and computational expertise and unfamiliarity with Bayesian methods on the part of policymakers. Researchers and stakeholders need advanced technical understanding to be able to discern the relative quality of different Bayesian methods. Few such individuals exist, creating the risk that conclusions from Bayesian analyses will be misrepresented to key stakeholders and misapplied in policy decisions. These shortcomings in adoption of Bayesian methods were echoed by members of the MEDCAC at the June meeting during which the MEDCAC members expressed a high level of confidence in Bayesian methods, but they highlighted the intensity of training that would be required to familiarize the future and current workforce of clinicians, policymakers, and others with the methods. These issues should be investigated more fully to ensure that an appropriate infrastructure and workforce exists to support high quality Bayesian analysis and interpretation. Without adequate infrastructure and resources, the potential exists for poorly designed, misrepresented, or misinterpreted studies to serve as the basis for policy decisions. Researchers and analysts at CMS and organizations conducting research need to understand the strengths and weaknesses of Bayesian methods, and how and why there are differences in studies using classical or Bayesian analyses.	We thank the reviewer for their thoughtful comment and will pass it on to CMS for their consideration.

Jose Ma. J. Alvir, DrPH	Pfizer, Inc	Implemen-tation	We agree that the articles by Sheingold [Shiengold, S. “Can Bayesian Methods Make Data and Analyses More Relevant to Decision Makers?” International Journal of Technology Assessment in Health Care. 2001; 17(1):114-22.] and Winkler [Winkler, R. “Why Bayesian Analysis Hasn’t Caught on in Health Care Decision-Making.” International Journal of Technology Assessment in Health Care. 2001; 17(1):56-66.] referenced in the TA provide several helpful suggestions for making Bayesian methods more accessible, such as more training materials and software, established test cases for using Bayesian analysis, and clear demonstrations of the value of this type of analysis in healthcare decision-making. We encourage both AHRQ and CMS to adopt policies and programmatic support to address the gaps in analytical skill and understanding and to prepare the workforce to handle studies with Bayesian methods.	We thank the reviewer for their thoughtful comment and will pass it on to CMS for their consideration.

Jose Ma. J. Alvir, DrPH	Pfizer, Inc	Implementation, continued	Use of Bayesian methods in CMS decision-making should be transparent and well-defined. We suggest that CMS' use of Bayesian studies include two types of transparency. The first layer of transparency should occur in the processes CMS uses to review and evaluate Bayesian studies. CMS should involve a range of relevant stakeholders (e.g., methodologists, clinical and other scientific experts, and patient and caregiver representatives) in the evolving discussion around use and interpretation of Bayesian methods in decision-making, to capture all viewpoints. CMS then should be transparent in defining instances in which Bayesian approaches will be considered and used in decision-making. CMS should also outline the criteria to assess Bayesian studies and be transparent in communicating to stakeholders about the decision-making process. The second layer of transparency should be at the research study level. By their nature, Bayesian methods afford transparency into research study design. Given the acknowledged subjective nature of inputs into Bayesian analyses, researchers must prospectively design studies and specify the prior distribution, which requires planning and consideration of how the results will relate back to the study inputs. CMS should consider developing processes, including communication strategies, to make the methods from studies used in decision-making as transparent and accessible to outside stakeholders as possible. For example, CMS could explore options to standardize reporting of methods and results for Bayesian analyses submitted for use in decision-making and assure that these reports are publicly available.	We thank the reviewer for their thoughtful comment and will pass it on to CMS for their consideration.

Jose Ma. J. Alvir, DrPH	Pfizer, Inc	Implementation, continued	CMS should build on other agencies' experiences with Bayesian methods. Future discussions should include an assessment of how CMS can build on the FDA's efforts to implement Bayesian methods. The FDA has worked with Bayesian methods for several years. The agency's 2006 Draft Guidance on the Use of Bayesian Statistics in Medical Device Clinical Trials helped to define parameters on the use of Bayesian methods specifically for devices. In adopting Bayesian analysis, we support efforts by CMS to adopt best practices and consider lessons learned from other agencies' experiences. The FDA has also recognized the need for education on Bayesian methods, both internally and externally. In recent presentations, Dr. Greg Campbell, Director of the FDA's Division of Biostatistics at the Center for Device and Radiological Health, has spoken about FDA's implementation of Bayesian methods in trials. His remarks highlighted the educational efforts the agency instituted, with internal courses and seminars to educate staff members and public forums to discuss Bayesian methods. [Campbell, G. "Bayesian Statistics at the FDA: The Pioneering Experience with Medical Devices." Presented at Florida State University Conference "Statistics, the Next 50 Years" on April 17, 2009. Accessed at http://www.stat.fsu.edu/Campbell.ppt on July 6, 2009.] CMS will need to adopt similar programs to educate staff members who will be analyzing Bayesian trial data.	We thank the reviewer for their thoughtful comment and will pass it on to CMS for their consideration.
Jose Ma. J. Alvir, DrPH	Pfizer, Inc	Conclusion	Pfizer appreciates the opportunity to provide AHRQ and CMS with comments on the draft TA on the use of Bayesian methods in randomized clinical trials. This report should serve as the beginning of a larger discussion on how to implement Bayesian methods and ensure that they are used appropriately. Pfizer welcomes any opportunity to discuss our comments and recommendations in further detail. Please feel free to contact me at 212-733-2051 with any questions, or if you need additional information on our above comments.	We thank the reviewer for their thoughtful comments.

Richard Chapell	Merck & Co., Inc.	General	Thank you for allowing us the opportunity to comment on the draft document. We have reviewed it thoroughly and believe that it describes a viable way forward for the use of Bayesian techniques. We especially applaud the way in which the twin dangers of Type 1 and Type 2 errors are highlighted in the discussion of subgroup analysis. We hope that the more methodical approach described in the document will come to replace the data-mining techniques that are commonly used at present.	We thank the reviewer for their comment.
Richard Chapell	Merck & Co., Inc.	Pg 8, ln 5	"combine" should be "combines"	The suggested change has been made (page 5).
Richard Chapell	Merck & Co., Inc.	Pg 9, ln 1	"where" should be "at which"	The suggested change has been made (page 5).
Richard Chapell	Merck & Co., Inc.	Pg 9, ln 6	Numeral "777" should be removed	The deletion has been made (page 5).
Richard Chapell	Merck & Co., Inc.	Pg 9, ln 20	Please add a period to the end of the sentence.	A period has been inserted (page 6).
Richard Chapell	Merck & Co., Inc.	Pg 16, ln 17	"around" should be "centered at"	The suggested change has been made (page 9).
Richard Chapell	Merck & Co., Inc.	Pg 20, ln 3	Please remove extraneous period.	The deletion has been made (page 11).
Richard Chapell	Merck & Co., Inc.	Pg 22, ln 13	Please remove extraneous comma and capitalize "It".	The suggested changes have been made (page 13).
Richard Chapell	Merck & Co., Inc.	Pg 24, ln 18	Please remove space between the final word in the sentence and the period.	The deletion has been made (page 14).
Richard Chapell	Merck & Co., Inc.	Pg 54, ln 8	"a cost-effectiveness decision models" please remove either the initial "a" or the final "s"	We have removed the "a" as suggested (page 31).
Richard Chapell	Merck & Co., Inc.	Pg 62, ln 3	Please remove the apostrophe from "it's"	The apostrophe has been removed (page 35).

Richard Chapell	Merck & Co., Inc.	Pg 67, ln 7	"Many of these questions are hoped to be explored…" should be "It is to be hoped that many of these questions will be explored…". Begin a new sentence with "Others"	We have restructured the sentence as requested (page 38).
Richard Chapell	Merck & Co., Inc.	Pg 80, ln 12	"is" should be "are"	The suggested change has been made (page 45).
Richard Chapell	Merck & Co., Inc.	Pg 80, ln 18	"regarded" should be "considered" or "regarded as"	The suggested change ("considered") has been made (page 46).
Richard Chapell	Merck & Co., Inc.	Pg 91, ln 1-5	A block of text appears to be missing.	We have added in the missing text (page 51).
Christine Fletcher, MSc	Amgen Ltd	General	I do not agree with some of the general points but there are schools of thought that would agree with the position: For instance the idea of taking epidemiological data and treating it as a prior to clinical data by weighting so as to "have the same weight" as the clinical data. David Spiegelhalter points out that adding a fixed amount to the variability in the prior is better, rather than increasing variance by a proportional amount seems.	We believe that the reviewer's comment is referring to the last paragraph on page 53 of the draft report (Case 2: Dissimilar Information). We have modified this paragraph to indicate that the weight assigned to prior information and the clinical data can be given more, less, or equal importance. We also now reference Spiegelhalter's work within this paragraph (pages 29-30 of final report).
Christine Fletcher, MSc	Amgen Ltd	General	Also: Bayesian methods seem to be being used because they are trendy. There is confusion between random effects hierarchical models and Bayesian methods, and which is the aspect that allows us to do certain things.	We agree with the reviewer and now note the first time that Bayesian hierarchical models are introduced (page 45) that hierarchical models are not limited to the Bayesian paradigm but are particularly natural within that way of thinking.

Christine Fletcher, MSc	Amgen Ltd	General	Could further clarification be given to the following points: Direct comparisons of meta-analyses between frequentist and Bayesian approaches (e.g., Bloom et al. 34) do not always yield consistent results – in particular, sometimes the results of the two approaches are similar and sometimes they are different. [If they are different it is crucially important to understand why]. However, some observations do appear to be reasonably consistent.	In general, in situations where extensive data are available, it is less likely that the conclusions of Bayesian and frequentist analyses will differ substantially. The bulleted points listed on page 30 detail additional situations which might result in either similar or different findings from meta-analytic approaches.
Christine Fletcher, MSc	Amgen Ltd	General	Estimates of efficacy from random-effect models have less precision than estimates of efficacy from fixed-effect models. This is nothing to do with Bayesian/Frequentist, but due to one model allowing for an extra level of variation (study to study) which is ignored in the other case. Both can be done within a frequentist paradigm.	We now include a discussion of fixed- and random-effects models (under both frequentist and Bayesian approaches) in a new section of the Tutorial ("Technical Note on Fixed- and Random-Effects Models, Heterogeneity, and Interaction"), page 12.
Christine Fletcher, MSc	Amgen Ltd	General	Fixed-effect models give greater weight to larger studies than do random-effects models. This will be mis-read. It is technically correct as a relative statement (with or without random effects), but it will be read in absolute terms ("fixed are better as they give more weight to this important studies"). What is true is that if there is heterogeneity between studies then we need to allow for that heterogeneity by downweighting the big studies and taking more account of the spectrum of different values across the studies.	See response to immediately preceding comment. We now include a new discussion of fixed- and random-effects models (under both frequentist and Bayesian approaches), heterogeneity, and interaction on page12.
Christine Fletcher, MSc	Amgen Ltd	General	Both approaches struggle a bit when the number of studies is small to moderate. In the fixed-effect model, this is reflected by a test for heterogeneity that has low power. In the random-effects models, this is reflected by the tendency for the results to be sensitive to the estimate (model B) or assumptions (model c) about ?. What they are saying is that with small numbers of studies it is difficult to estimate the variability from study to study. As such using methods that work well with small amounts of data in random effect models are important (e.g., the KR adjustment).	We agree with the reviewer that both approaches struggle a bit when the number of studies is small to moderate, and this is reflected within the text (page 30, bullet point 3).

Christine Fletcher, MSc	Amgen Ltd	General	The results of the fully Bayesian analysis are most likely to differ from others when relatively little information is available from the data. This is, in general, the most dangerous circumstance for drawing definitive conclusions – which phenomenon should be illustrated by a careful sensitivity analysis. Correct! The most promising circumstance to apply a fully Bayesian approach occurs when the type of information available to the analyst is sufficiently disparate as to call into question the other two models. I think this means when there is heterogeneity between studies. If so, then I do not agree.	We agree with the reviewer about the most dangerous circumstance for drawing conclusions about a Bayesian analysis, and also about those circumstances in which a Bayesian analysis is promising. To clarify that we were not discussing the heterogeneity between studies in the last bullet on page 54 of the draft report, we have removed this bullet concerning the "most promising" circumstance and instead now describe how a situation where RCT data are modest and external information is available allows for a particularly natural application of Bayesian techniques (page 30, last bullet point).
Christine Fletcher, MSc	Amgen Ltd	General	From attending several meetings with the academic health economists driving the HTA process within NICE, it's clear that the Bayesian view is not only dominant, it is almost unanimously held, with some individuals this goes to the point at which they don't even see a Bayesian/Classical debate anymore. I know this is anecdotal, but we (industry) have got to get the same level of organizational cognizance about the detailed techniques as we currently have with the, largely classical, regulatory requirements.	The reviewer's comment is noted, and we agree that the acknowledgement of potential applications of Bayesian methods is becoming more widespread within certain communities. However, from our own experience, the use of Bayesian methods is still not widely understood and applied. This unfamiliarity led CMS/AHRQ and the Duke EPC to work on this report.

Christine Fletcher, MSc	Amgen Ltd	General	I don't like the debate being cast as Bayesian/Frequentist. I see frequentism as an approach to probability, not inference. It is perfectly possible for both prior and posterior functions to be frequency based. Indeed I got a Bayesian to admit to this once when he confirmed that, if agreement about a prior could not be found, a vote could be taken – which is about as frequentist as you can get! I prefer the term "Classical."	In this report we refer to Bayesian versus frequentist approaches where the common term "frequentist" encompasses "classical" approaches. The essential interchangeability of these two terms for the purpose of this report is indicated on page 5.
Christine Fletcher, MSc	Amgen Ltd	General	The distinction between classical multi-level error models and Bayesian models needs to be well defined. I have, again anecdotally, seen people miss the difference between the two.	We agree with the reviewer and now note the first time that Bayesian hierarchical models are introduced (page 45) that hierarchical models are not limited to the Bayesian paradigm but are particularly natural within that way of thinking.
Christine Fletcher, MSc	Amgen Ltd	General	There is full support on the conclusions and recommendations regarding subgroups, and the report should be commended for tackling a variety of current statistical issues pertaining to areas surrounding evidence synthesis, the methodological considerations for incorporating randomized and observational data, and the specific issues dealing with subgroup analyses. The report includes practical examples and the simulations appropriately investigate methodological aspects analysts have to deal with in this area.	We thank the reviewer for their comment.
Karen Lynn Price, PhD	Eli Lilly and Co.	General	Expert opinion – when considering an expert opinion, it is prudent to be wary of potential biases that are related to expert opinions (too enthusiastic, span too narrow of the parameter space) – we recommend that the use of Bayesian methods to assess benefit/risk be considered. For example, Bayesian methods can be used to calculate the probability that, for example, the benefit exceeds the risk.	We now remind the reader of the importance of sensitivity analyses on the prior distributions based on expert opinion (page 20). Although this is important, because it is not the focus of our discussion, and in order to preserve clarity, we have not addressed the use of Bayesian methods to assess benefits and risks.

Karen Lynn Price, PhD	Eli Lilly and Co.	Pg 52	Any suggestions regarding what is considered small to moderate? (3rd bullet)	Just as with sample size calculations, what is considered small to moderate would depend on the variability of the outcome of interest within and between trials and goals of analysis. We now clarify this on page 29.
Karen Lynn Price, PhD	Eli Lilly and Co.	Pg 90-91	It appears that the thought was not finished.	We have clarified this sentence (page 51).
Karen Lynn Price, PhD	Eli Lilly and Co.	Pg 90-91	There should be some mention of what to do if one trial appears aberrant.	If a trial appears aberrant, one may adopt a cross-validation approach, considering the analysis with and without the data arising from that particular trial. This would allow us to assess how influential the trial might be in the overall conclusions. We now add this explanation into the text on page 51.

1 Names are alphabetized by last name. Those who did not disclose name are labeled "Anonymous Reviewer 1," "Anonymous Reviewer 2," etc.

[2] Affiliation is labeled "NA" for those who did not disclose affiliation.

[3] Page and line numbers refer to the draft report.

[4] Page and line numbers refer to the final report.