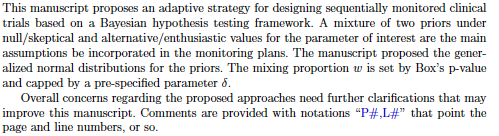
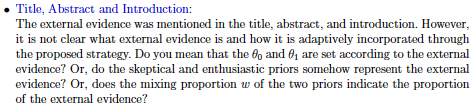
\*Reviewers' Comments to Authors\*

\*Reviewer 1\*

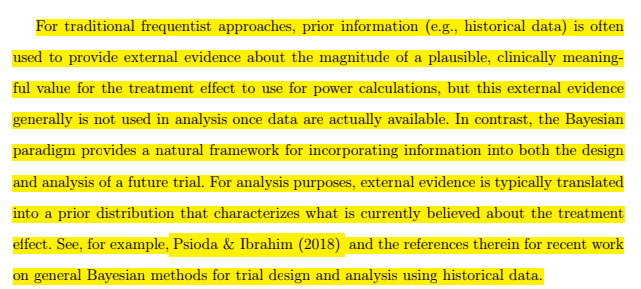
Comments to the Authors:



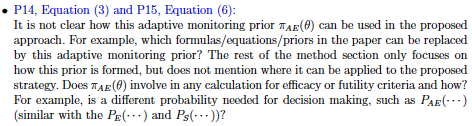
**[Comment 1]**



**[Response by Authors]** The authors added the paragraph below to the introduction to describe how external evidence is incorporated through the prior distributions used for analysis. Based on our framework, is set according to external evidence, the enthusiastic monitoring prior represents external evidence about the treatment effect, and the mixing proportion represents the applicability of external evidence in the analysis. Typically, represents a null treatment effect that is not informed by external evidence.

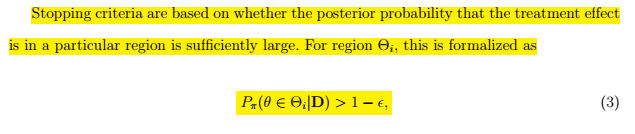


**[Comment 2]**

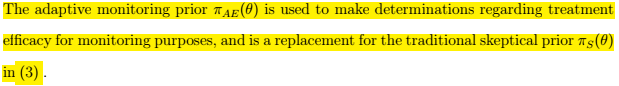


**[Response by Authors]**

To show how the monitoring prior, is used in the proposed approach, the stopping criteria was given its own line in Equation (3) of Section 2.1.4 “Maximum Sample Size and Formal Stopping Criteria,” and was generally defined with respect to any prior distribution:



The following sentence was added to Section 2.2.3 “Incorporating Prior Information in the Monitoring Priors” to show what calculation uses the adaptive monitoring prior:



Equivalently stated, the probability for decisions about efficacy is replaced with .

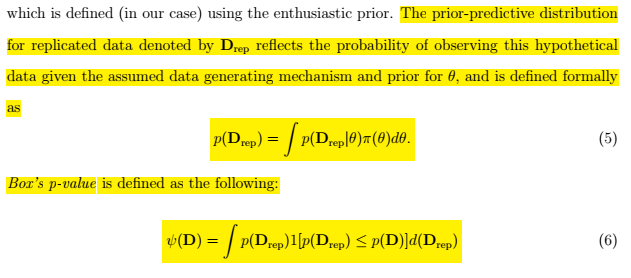
[**Comment 3]** ****



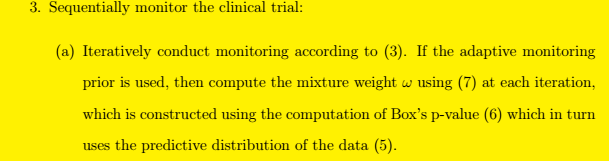
**[Response by Authors]**

Yes, as originally formulated, the adaptive monitoring prior uses *Dobs* in computing the mixing weight given to the enthusiastic prior component Equation (4), and in the assessment of the stopping criteria Equation (5). Note that the new Equation (3) introduced in response to Reviewer #1 [Comment 2] has shifted the original equation numbers one number higher.

The authors appreciated the chance to clarify the use of *Dobs* and *D*, and it was correctly pointed out that *Dobs* would be equivalent to *D* in that instance. This presented an opportunity to rectify the ambiguous use of notation throughout the entire manuscript. The term *Dobs* was dropped completely; instead, *D* always refers to observed data and the term *Drep*refers to hypothetical data used in the expression for the predictive distribution and Box’s p-value. The revised Equations (5) and (6) are below:



This comment, in addition to Reviewer #2 [Comments 3] led to the creation of Appendix C “Step-by-Step Implementation Guide” to address how exactly the formulas used can be applied in a trial. This includes references to how these equations are utilized in a trial:



**[Comment 4]**

**[Response by Authors]**

The authors agree that this is a vague expression. To make this statement consistent with the existing notation for the mode of a probability distribution first used in Section 2.1.3, the following change was made:



\*Reviewer 2\*

Comments to the Authors:

*This article provides a Bayesian framework for sequential monitoring of clinical trials using external data.  It defines skeptical and enthusiastic priors in the context of Bayesian hypothesis testing and proposes a two-component mixture prior to combine these two types of priors. Weight of each component is determined based on an assessment of prior data conflict, for example, a higher weightage will be given the skeptical component if the observed data are not compatible with the enthusiastic prior. Section 2 of the article presents a thorough discussion on the elicitation of prior and a computation strategy to determine the weights.*

*Overall, the article is interesting but there are scopes of further improvement.  My comments are provided below:*

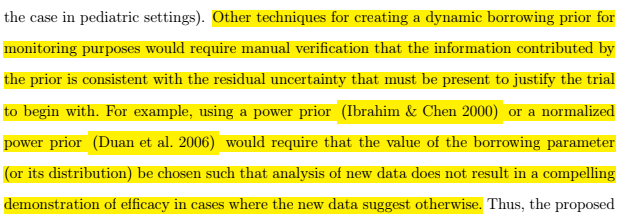
**[Comment 1]**

*Authors provide a very limited discussion on the existing Bayesian approaches for sequential monitoring of clinical trials using noninformative prior and other approaches of elicitation of informative prior (such as power prior). It will benefit the readers if the authors discuss how their proposed method differs from other existing approaches through examples/simulation study.*

**[Response by Authors]**

A comparison of our method with a Bayesian approach using a noninformative prior is now given in the new Section 3.3 “Comparison to Single Analysis with Non-Informative Prior.” (See Reviewer #2 [Comment 2]).

A reference comparing a feature of our method to other approaches of eliciting an informative prior is given in the discussion:

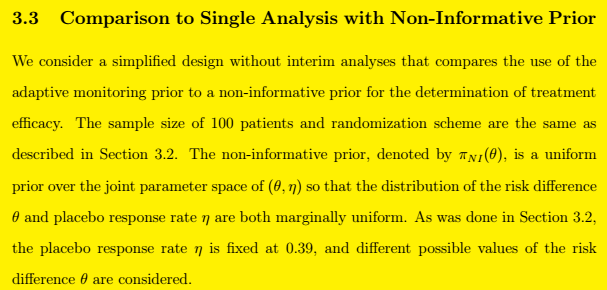
****

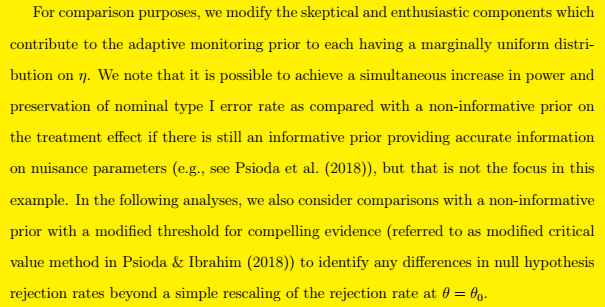
Existing Bayesian approaches to sequential monitoring are discussed in the introduction with cited works including Spiegelhalter et al. 1993 and Jennison & Turnbull 2000. Comparison to existing Bayesian approaches to sequential monitoring is provided by using the default skeptical prior for determination of treatment efficacy in Section 3.1 “Single-Arm Trial with Binary Endpoint.” Articles cited that discuss operating characteristics of existing Bayesian approaches include Psioda & Ibrahim 2018, Kopp-Schneider et al. 2020, etc.

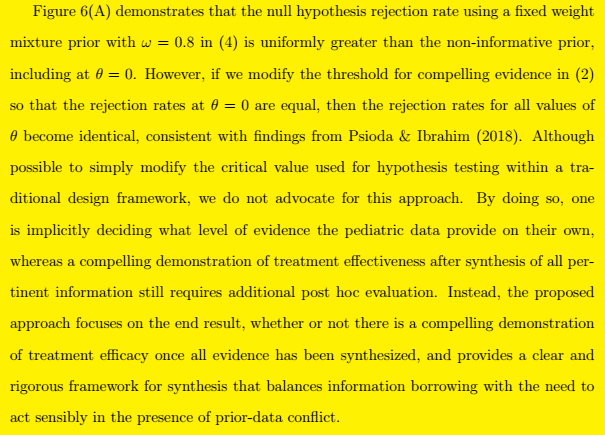
**[Comment 2]**

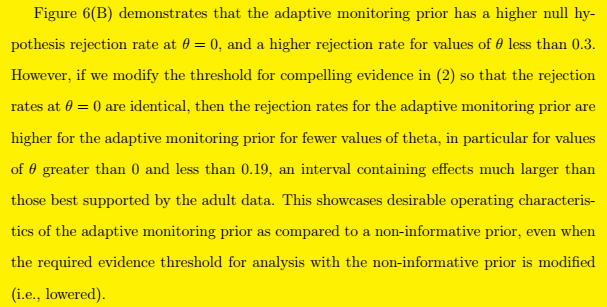
*This paper notes that strict type 1 error control may not be achievable when prior information is incorporated into the analysis. I think it will be useful for the readers if the authors provide a more detailed discussion on this topic. I recommend the authors to consider performing a simulation study, comparing the operating characteristics based on a standard design with noninformative prior and the proposed design using the mixture prior.*

**[Response by Authors]** The authors provide a detailed discussion of this topic along with results from the suggested simulation study in the new Section 3.3 “Comparison to Single Analysis with Non-Informative Prior.”







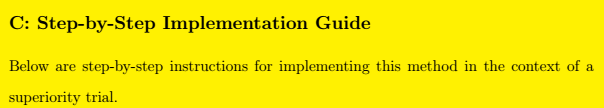


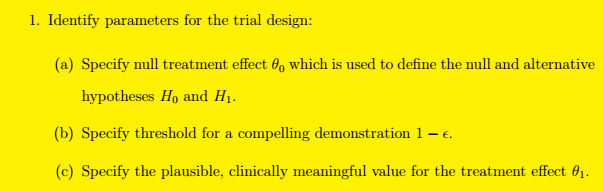
**[Comment 3]**

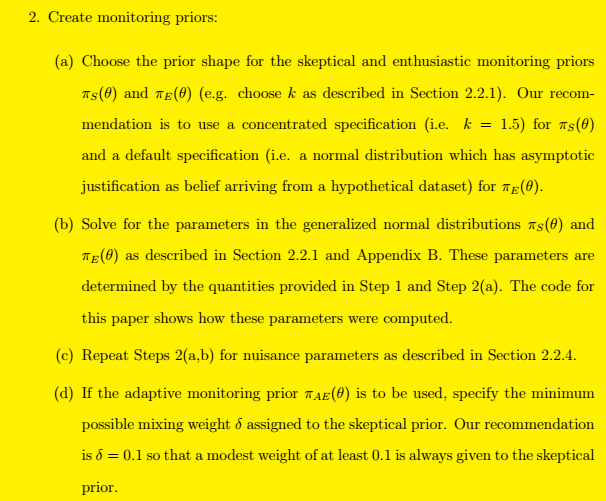
*I recommend the authors to provide a step-by-step algorithm for implementing their approach. This will be useful for the practitioners.*

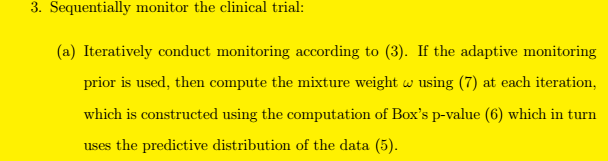
**[Response by Authors]**

The authors added Appendix C “Step-by-Step Implementation Guide” to address this comment.









**[Comment 4]**

*In my opinion, Section 5 of the paper can be improved by the authors taking a more critical view of their writing, sharpening the arguments, and including recommendations for the practitioners.*

**[Response by Authors]** The authors removed text from the discussion that was redundant with previous exposition, and focused the discussion on two main conclusions: the requirement that observed data must demonstrate some degree of efficacy on their own to justify stopping enrollment early, and a comparison with a published post-hoc Bayesian analysis using a hierarchical model. A reference to Appendix C which includes an implementation guide for practitioners was included.

In taking a more critical view of our writing, we removed statements that were vague or not fully supported by the results of the paper, such as the following sentence referring to concepts such as “overwhelming treatment benefit,” “in the more likely scenario,” “some evidence of benefit,” and “reasonable compatibility” was removed since those concepts were not defined rigorously:

