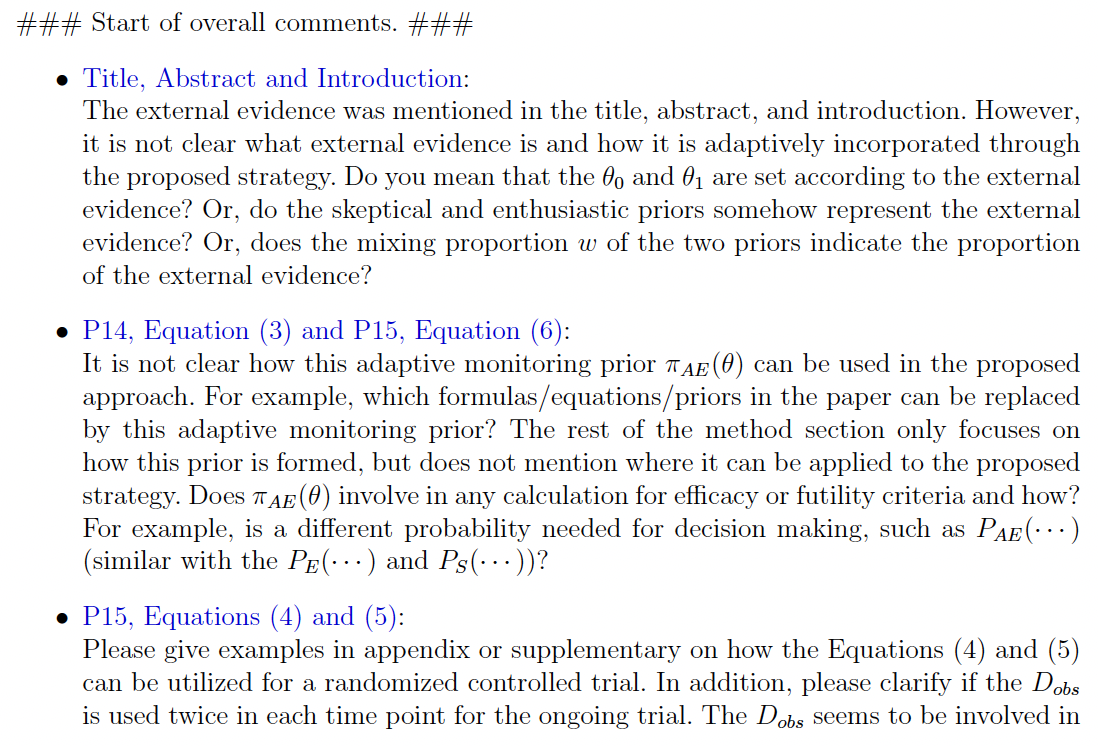
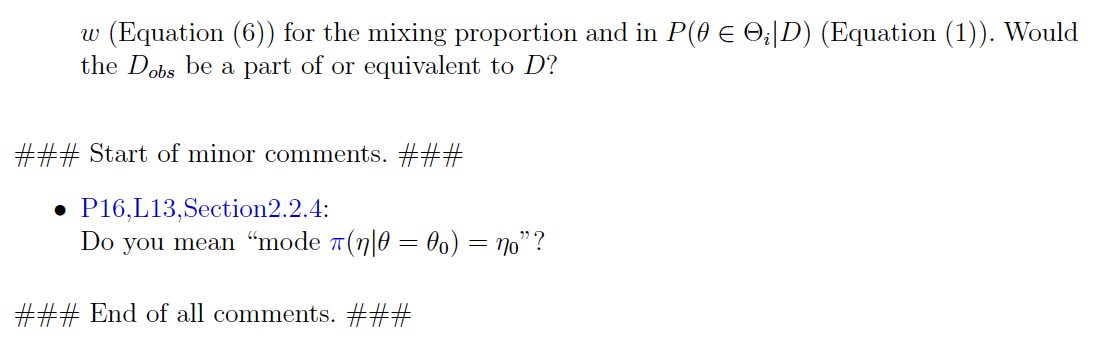
# Reviewer 1





1. Yes regarding \theta\_1. Could make last sentence of paragraph 2 from introduction into a new paragraph 3.
   1. Information vs. evidence
   2. Evidence points to probabilistic statements regarding efficacy
      1. Step 1: External information just to get \theta\_1 (effect of interest) (e.g. frequentist power calculation) (design)
      2. Step 2: External evidence once becomes prior (analysis)
2. At the bottom of page 15 we can revisit equation (2) to show how this is used exactly.
   1. Replace >= in (2) with = (check second line from the bottom on pg8, should have been equal all along)
   2. Explain how posterior probability of efficacy (design rule) can use \pi\_S, \pi\_E, \pi\_AE are used (break out to get equation #)
3. We can clarify how (5) is used precisely, (i.e. Table 2 and Figure 6). I’m not sure how to demonstrate (4).
4. The distinction is between the mode of a r.v. and the mode of the density of a r.v.

# Reviewer 2

Major comments

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.
   * Only way to get reasonable T1E control using NI prior is to modify threshold, which changes concept of compelling evidence (modify critical value)
   * Changes posterior probability from being quantification of evidence
   * Contrast idea of having quantification of substantial evidence not changing based on # of looks
   * Our monitoring process is different since we are using external evidence into the monitoring prior, is new and different from existing sequential monitoring approaches
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.
   * No sequential monitoring, one look (standard design = one look)
   * Calculate power for standard design with NI prior and standard evidence threshold (e.g. SAS proc power)
   * Compare our \pi\_AE power/T1E
     1. Should be same power (no free lunch)
     2. Allow T1E for NI prior to be as inflated as ours is (modify critical value). Reference Psioda 2018 Biostatistics paper, power is identical across the board. Note loss of connection to level of substantial evidence.
3. I recommend the authors to provide a step-by-step algorithm for implementing their approach. This will be useful for the practitioners.

Minor comment

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.

1. (MS) Reviewer 2 Comment 1: I am by no mean a card carrying Bayesian, but it seems that this comment is not completely aligned with the “philosophy” of the proposed method. Basically that philosophy is that when monitoring a trial and wanting to make decisions on accruing data what do you need to do to (a) convince a skeptic that the results are actually favorable to stop, (b) convince an enthusiast that the results are unfavorable to stop, or (c) if you don’t stop what is a reasonable estimate of the effect and its uncertainty. If one has to multiple ways of soliciting the priors it seems this would be exhaustive for a paper.

(EK) I agree with your response to Comment 1. We should address how this method differs from existing methods through discussion of the literature and without relying on simulation studies.

1. (MS) Reviewer 2 comment 2: I believe this understanding about inflation of the Type I error when using an informative prior is published. Rather than simulation, is it possible to reference the literature here?

(EK) Furthermore, regarding Comments 1 & 2 on simulation studies: comparing to "existing Bayesian approaches for sequential monitoring of clinical trials using noninformative priors" is tricky since both the priors and the decision rules are involved in the designs. If we simply replaced our mixture prior with a noninformative prior while still using our decision rule for efficacy (i.e. our significance threshold for efficacy), then the noninformative prior would result in conclusions of treatment efficacy very often since it is much less skeptical than our version of the monitoring prior for efficacy. If we are comparing designs with different significance thresholds, then this could lose sight of our fundamental philosophy of the proposed method (as Mat stated). Also, this could expand to the wider discussion of comparing Bayesian methods with frequentist group-sequential designs which is well established in the literature, and could further distract from our method.

I'll discuss with Matt tomorrow to see how we can comprehensively address the requests for simulation studies through discussion and where some simulated results may be necessary. Perhaps we can add a couple well-placed examples of analyses that use different priors at points in our discussion to satisfy the reviewer. ~~For example, we mention that some priors which don't maintain residual uncertainty that the alternative hypothesis is true could result in conclusions of treatment efficacy even with a negative trial result (i.e. adult data could completely overwhelm the pediatric trial result). Perhaps we could show empirically how this could happen with our dataset and a power prior that does not maintain the residual uncertainty.~~

1. Add to Appendix C
2. Make own paragraph in discussion