We thank the reviewers for the time and expertise they have invested in these reviews. Below are our responses.

Response to Reviewer 1

Comments on paper  
------------------  
  
1. The examples and specifically the interpretation of the parameters could be expanded to be a bit more "stand-alone". When I am searching for a package to use I often go straight to the examples, so that would help people like me.

**Answer:**

**We have added the interpretation of all parameters in the Examples section (see page 10). We have also included a linked reference (see page 10) to the definition of Bayesian power and type I error where it appeared in the manuscript.**   
  
Comments on package  
-------------------  
  
1. The user-facing functions all return lists. It would improve the user experience a lot, but not be much burden to instead return S3 objects with print and/or summary methods to display only the most relevant information.

**Answer:**

**We have** **turned all the return values into S3 objects and implemented *summary* methods which display the most relevant information returned by each function.**   
  
- No automated testing apparent as part of the package other than the examples

**Answer:**

**We have added automated testing of some of the functions in the package using the “testthat” R package. Specifically, we test that the power.two.grp.fixed.a0() function produces similar power and type I error calculations compared to Table 2 in Chen et al. (2011) (absolute difference < 0.01). We also test that the power.two.grp.random.a0() function produces similar power and type I error calculations (absolute difference < 0.02) with a beta(1000, 1000) prior on a0 compared to the power.two.grp.fixed.a0() function when a0 = 0.5. We also test that the glm.fixed.a0() function produces similar posterior mean estimates of beta (all absolute differences < 0.02) compared to the results produced by PROC MCMC on SAS.**

- Documentation is good, but there is no vignette, no webpage, no NEWS file, and no README. I'd encourage you to take some of the content from this paper and include a vignette/more extensive readme.

**Answer:**

**We have added a vignette demonstrating Bayesian sample size determination for two group models with binary and normal outcomes. We have also added a NEWS file documenting the recent updates to the package.**

- The code itself is easy to read and commented appropriately and in a useful way.

**Answer:**

**We agree and thank the reviewer.**

Response to Reviewer 2

C1:  Computation of the normalizing constant c(a0) requires integrating over a region with the same dimension as θ. The computation of c(a0) seems to work perfectly fine for the dimensions in the types of examples discussed in the paper. At which dimension of θ does the computational strategy for c(a0) used in the package begin to run into problems? It might be useful to mention this so users can be aware of this if they want to try higher dimensional problems.

**Answer:**

**The following text has been added to the paper (page 4) and the documentation of the package:**

**“In the Examples section below, we demonstrate computing the normalizing constant for one historical dataset with three covariates. Due to computational intensity, the normalizing.constant function has not been evaluated for accuracy for high dimensional β (e.g., dimension > 10) or high dimensional a0 (e.g., dimension > 5).”**

The normalizing constant in the computation of c(a0) is designed to ensure that R2 > 0.99. For higher dimensional problems, is it possible that one cannot satisfy R2 > 0.99. If so, is an error message printed out to the user with a recommendation on how to resolve the issue?

**Answer:**

**There are a few scenarios in which R^2 > 0.99 cannot be satisfied. If** **R^2 > 0.99 cannot be satisfied, the following error message is printed to assist the user:**

**“Some coefficients are not defined because of singularities. Potential causes include number of grid rows too large or too small, insufficient sample size of the historical data, insufficient number of iterations for the slice sampler, or near-zero grid values.”**

C2:  For functions such as power.two.grp.fixed.a0, the output returned to the user is rather sparse. For power.two.grp.fixed.a0, it just returns the estimated power. However, many users might find additional posterior output to be useful. For example, in the power.two.grp.fixed.a0, one could return the draws of P (μt − μc < δ|y(n),π(f)), and perhaps samples y(b) ∼ f(y(b)|θ(b)). A similar comment could be made about the power.two.grp.random.a0 function.

**Answer:**

**For power.two.grp.fixed.a0() and power.two.grp.random.a0(), we have added to the output the posterior probabilities of the alternative hypothesis, the average posterior mean of mu\_t and mu\_c and their respective bias. For power.glm.fixed.a0() and power.glm.random.a0(), we have added to the output the posterior probabilities of the alternative hypothesis, the average posterior mean of beta and their respective bias. We have also** **turned all the return values into S3 objects and implemented *summary* methods which display the most relevant information returned by each function. We do not return draws of the data y(b) as the resulting matrix would be quite large if the number of simulated trials is large.**

C3:  It appears that the “glm” functions (glm.fixed.a0 and glm.random.a0) require the “current” data to have a treatment column in the design matrix. From the description of power priors for GLMs (in the Sections before the “Using BayesPPD” section), the power prior description appears to focus on the case where the historical data and the current data share the exact same set of covariates. Hence, I was a little bit confused when I first attempted the glm-based functions in BayesPPD. It might be helpful if the earlier discussion on GLMs can mention that the current dataset will have an additional treatment column.

**Answer:**

**The following text has been added to page 3 of the paper where the GLM is defined:**

**“By default, BayesPPD assumes the historical data consists of control group subjects only. Therefore, the historical covariate matrix does not have the treatment indicator variable, while the current covariate matrix does. The package also allows the historical data to be used to inform the treatment effect parameter; then the historical and current covariate matrices will both have the treatment indicator.”**

Also, not allowing a treatment column in a historical dataset seems to make things 1

less general than they could be. Is there any reason why one could not perform a power calculation while incorporating partial information about the treatment effect from historical data?

**Answer:**

**We appreciate the suggestion and have added the feature to the package. By default, the package assumes the historical data is composed of control group subjects only. If the user wants to use the historical data to inform the treatment effect, one can set borrow.treat=TRUE and include the treatment indicator in the historical covariate matrix. The paper has been edited to reflect this change.**