

We thank the Editors and Reviewers for their helpful comments and suggestions. In response to the specific questions:

**Reviewer 1:**

**In addition to the treatment with thiopurines, is there any other treatment used to the patients?**

In the revised manuscript we have included a supplementary table with a summary of other treatments used in the thiopurine-exposed patients (Suppl table 3).

**Pregnant women with active IBD have higher neonatal LBW and SGI. These women were also treated with thiopurine. How to distinguish the outcome of higher LBW and SGI caused by active IBD or by thiopurine? In treatment & LBW or treatment & SGA, the cases in the non-thiopurine exposure not only included the IBD active cases, but also included IBD non-active cases. As this way, it is hard to compare the effect on LBW/SGA were only by thiopurine or activity of IBD. If the authors can compare the LBW/SGA between active IBD and active IBD with thiopurine treatment, the result will be more powerful to demonstrate if thiopurine has side effects on fetal growth.**

The Reviewer correctly points out that the majority of studies reporting an impact of thiopurines on birth outcomes did not adjust for disease activity. The one study that did adjust for this found no impact of thiopurines on these birth outcomes. These findings are noted in the manuscript (Results, Page 12);

“There was a higher risk ratio of LBW in women on thiopurines (RR 1.4; 95% CI 1.1-1.9; p=0.007, Figure 5), however, when we only included the study that adjusted for disease activity were included, there was no significant effect of thiopurines on LBW (RR 1.2, 95% CI 0.6, 2.2; p=0.6) [19]. “

**Figure 1: the figure is not clear.**

We have provided an updated Figure to improve its clarity.

**In the section of Results, the last paragraph, please describe more detail for the sentence: when we only included the study that adjusted for disease activity were included, there was no significant effect of thiopurines on LBW.**

This sentence has been edited for clarity in the revised manuscript (Results, Page 12);

“There was a higher risk ratio of LBW in women on thiopurines (RR 1.4; 95% CI 1.1-1.9; p=0.007, Figure 5), however, when we only included the study that adjusted for disease activity were included, there was no significant effect of thiopurines on LBW (RR 1.2, 95% CI 0.6, 2.2; p=0.6) [19].“

**The duration of IBD activity and exposure to thiopurine is not defined in the study. The reviewer consider this is a important factor related to the effects to outcomes of pregnancy.**

In only one study (19), patients were grouped by timing of their flare-up (early pregnancy, late pregnancy or both). They reported a 3-4-fold risk increase of low birth weight if the flare event was in both early and late pregnancy. No other study has examined birth outcomes in this manner to date. This data has been added to the revised manuscript (Results, Page 11).

### **Reviewer 2:**

**I see from the text that the authors mention the "inclusion criteria was not otherwise restricted by study size, language or publication type". Does this mean that you evaluated studies in other languages as well? If yes, how did you do this as there might have been several different languages other than English. Please clarify or revise.**

We identified all abstracts that had an English-language summary, which is standard for PubMed, EmBase and Thompson Index. If any paper from this search was in a language other than English, we would obtain a translation of that paper into English for analysis. In this case, no abstracts were identified where the language of the primary paper was in another language. This is a limitation of the search in papers only had a record of the abstract in their local language. This has been clarified in the Methods (Page 8);

"Only abstracts summarized in English were used for screening purposes."

**The authors mention that "There was not a homogeneous definition of 'active disease' in the included studies". I believe it would be interesting and also important to the readers to know which definitions were used in different studies. The authors may put the definitions and related studies in a table.**

We have now added a new table in the revised manuscript with the criteria for "active" used in each study (Table 2).

**Table 1 needs amendment and following information needs to be in the table: Study ID, Sample size, Study design/recruitment/study period, number of patients with/without active disease, outcomes.**

Table 1 has been updated in the revised manuscript.

**Please put reference next to each study in the Table 1 and Supplementary table S2.**

**Supplementary Table S1 can be removed from the article as it does not add to the knowledge of the readers about the studied issues**

References have now been added to table 1 and Supplementary S2 to the effects to outcomes of pregnancy.

### **Reviewer 3:**

**However, subjects data were not adequately described in the manuscript, the authors should provide a table with demographical and clinical characteristics of pregnant women with IBD.**

Revised Table 1 now includes further data on the included subjects from the studies. Since there were different levels of detail on subjects in each study, we summarized those details that were universally reported.

**The source of the data is reliable; however, the authors should explain more specifically the reasons of exclusion of 22 studies**

The excluded studies are listed in Supplementary table 2. The "reasons for exclusion" are listed, according to standard Cochrane Collaboration convention (based on population, inclusion criteria and outcomes).

**The forest plot showed in figure 4 (RR of SGA based on Thiopurine exposure) should be implemented by other relevant studies. The analysis based on only two articles is not statistically powerful.**

We agree that meta-analysis with less than 5 studies lack generalizability to larger populations. However, for this outcome, only two studies met inclusion criteria and reported raw data in a manner suitable for meta-analysis. This is an acknowledged limitation of our study (Discussion, Page 14);

“For some outcomes, less than 5 studies were suitable for meta-analysis, which may limit the application of the results to the pregnant IBD population as a whole.”

**In addition, the results of sensitivity analysis are not clear.**

As noted in the Discussion, there were too few studies for sensitivity analyses;

“We intended attempted to performed sub-analyses for the case definition, (as reported in the Results) disease activity definitions, follow-up duration and protocol but there were too few similarly defined parameters within the included studies to further sub-group these for analyses.”

**We found two pertinent references using the keywords reported in the abstract that can be potentially useful for the manuscript. (“Association between maternal inflammatory bowel disease and adverse perinatal outcomes” Getahun D. 2014; “Pregnancy outcomes in women with inflammatory bowel disease--a population-based cohort study” Kornfeld D. 1997)**

These studies were both reviewed and are included in the “Excluded Studies” table (Supplementary Table 1).

**Language editing certificate:** We don’t provide a language certificate because Dr. AC Moss is a native speaker of English.