Format for ANSWERING REVIEWERS

September 23, 2015

Dear Editor,

Please find enclosed the edited manuscript in Word format (file name: 17863-Review.doc).

Title: Hypoallergenic formula with Lactobacillus rhamnosus GG for babies with colic: A pilot study of recruitment, retention, and fecal biomarkers

Author: Nicole Y Fatheree, Yuying Liu, Michael Ferris, Melissa Van Arsdall, Valarie McMurtry, Marcela Zozaya, Chunyan Cai, Mohammad H Rahbar, Manouchehr Hessabi, Ta Vu, Christine Wong, Juleen Min, Dat Q Tran, Fernando Navarro, Wallace Gleason, Sara Gonzalez, J Ma

Name of Journal: *World Journal of Gastrointestinal Pathophysiology*

ESPS Manuscript NO: 17863

The manuscript has been improved according to the suggestions of reviewers:

RESPONSES TO REVIEWERS:

Reviewer 29421:

1. "..It appears to have started out as a properly constructed randomised controlled trial with appropriate study size calculations that had to be abandoned for lack of recruitment within the funding timeframe. **The reviewer is correct, but many significant findings were made during this 2.5 year study, and important information can be used toward future RCT planning. Our institution believes that it is unethical to not publish work that has entailed substantial time and commitment on the part of infants and their parents.**

2. In relation to 1, above there is confusion within the manuscript as to what the a priori aim was. For instance under the section "Sample size and power" the authors say that the study aimed to determine recruitment, retention, adverse events and biomarkers but was not powered to detect differences in crying time between the two studies arms. **The study yielded important information on the above components that will be needed for a future RCT. The optimal time-points and biomarkers are enumerated clearly in the abstract, and the optimal sample size for a RCT is indeed 60. Note that the Consort Diagram shows that (at least in the U.S.), considerably more than 60 enrollees will be required, highly suggestive that a multicenter design is needed.**

3. While it is worthy of the authors (and of the peer review process) to extricate relevant information from an otherwise unsuccessful study I am not sure that this particular project yielded novel information over and above the difficulties of performing clinical trials among babies with regulatory disorders in particular. **Novel findings of our study were: (a) The optimal time point for potentially observing differences in crying and in fecal calprotectin were days 14 and 90, respectively. This may allow planning that does not require as many visits or stool exams as those which were done. (b) Microbial communities were chaotic in infants with colic, even more so than reported in Dutch infants by DeWeerth et al., 2013. (c) Our study was the first to analyze cytokine levels and circulating Tregs in infants with colic. (d) The gradual "crowding out" of L. rhamnosus after 14 days, despite continued consumption of LGG was unexpected, and (we believe)**

interesting.

4. There were no differences for any of the parameters examined between the two groups at any point in the study. Therefore it is not appropriate to include statements in the abstract or elsewhere which could be misconstrued as showing a difference in favour of the probiotic arm (In particular, "The maximal difference of crying + fussing time was observed at day 14, comparing the 2 groups, with a mean difference of -91 (95% CI: -76, 259) minutes, favoring the LGG+ group") nor indeed that calprotectin "showed a trend" towards a decrease in the probiotic arm. **These points are acknowledged and the MS was revised. It is fair to say, however, that data that are not significant in a pilot study can be used to plan future larger trials and help to determine optimal markers and time points.**

5. It is not clear why the particular circulating cytokines that were measured were chosen. **The panel includes "standard" pro-inflammatory cytokines, 3 of which were shown by Yan and Polk to induce apoptosis in intestinal epithelial cells. In those studies, LGG was protective (PMID 12393915).**

6. I would be concerned regarding the loss of blinding by the use of a sticky label to partially cover the formula containers. **The sticky label was covered with a can-encircling clear tape that made it virtually impossible to remove the label. Our pharmacists tested this technique before the study began.**

7. It would seem that the colonisation by the probiotic organism used did not persist (figure 3A). This is a potentially major issue suggesting, amongst other possibilities, that either the children were no longer getting the appropriate formula, that the organism was no longer viable within the formula or this particular probiotic rapidly becomes excluded by the normal microbiota in babies? **Colonization did not persist in the stool, although lactobacilli are felt to primarily colonize the small intestine. We previously showed in adult volunteers that a similar probiotic, L. reuteri, often could not be detected in the stools even when we were certain that the product was being taken, based on measurements of the amount of liquid probiotic remaining in the vial (Mangalat et al, PMID 22970150). However, despite low or undetectable levels in the stool, an antimicrobial peptide (calprotectin) was observed to significantly and uniformly rise in stools of volunteers taking the probiotic, and not in volunteers taking a placebo, indicating that it did have biological effects.**

Reviewer 34168

We modified the paper to follow the CONSORT 2010 checklist.

Reviewer 188382

1. In Table 2 there is no "p" value for the differences presented between LGG+ an LGG- group per visit. This is a key data that will demonstrate if LGG had or not a significant impact on crying + fussing time, and should be addressed.

Table 2: Longitudinal analysis of clinical variables by study group (LGG+ vs. LGG-) at baseline and follow up visits

	Adjusted means (95% Confidence Intervals (CI))			P-value
	LGG+ group	LGG- group	Mean differences(95% CI)	
Crying+fussing time (mins)				
Visit 1 (Baseline)	296 (210, 381)	337 (251, 422)	41 (-80, 161)	0.51
Visit 2	197 (117, 278)	289 (142, 436)	91 (-76, 259)	0.29
Visit 3	144 (54, 234)	199 (69, 328)	55 (-104, 213)	0.50
Visit 4	111 (65, 157)	133 (60, 205)	22 (-64, 107)	0.62
Fecal calprotectin (µg/g)				
Visit 1 (Baseline)	285 (199, 371)	294 (184, 404)	9 (-131, 149)	0.90
Visit 2	226 (182, 270)	305 (186, 423)	79 (-48, 205)	0.22
Visit 3	229 (113, 345)	250 (154, 347)	21 (-130, 172)	0.78
Visit 4	211 (80, 342)	332 (225, 440)	121 (-48, 291)	0.16

Longitudinal model: barr diary data=β_0+ β_1*visit2+ β_2*visit3+ β_3*visit4+ β_4*group+ β_5*visit2*group+ β_6*visit3*group+ β_7*visit4*group; Here, visit2, visit3, visit4 are dummy variables; visit2=1 if at visit 2, 0 otherwise; visit3=1 if at visit 3, 0 otherwise; visit4=1 if at visit 4, 0 otherwise; group=1 if in LGG group, 0 otherwise.

2. If differences on crying + fussing time between LGG-treated and non-treated infants during different visits was significant, but no significant changes in blood cytokine, stool microbiota, gas and calprotectin were detected, which pathophysiological hypothesis the authors will support to explain these results? **We believe that a larger sample size would flush out potential differences in crying time and fecal calprotectin; however, only 4 studies to date have been completed, and the numbers are still small. Moreover, 3 or the 4 studies looked only at breast-fed infants. If inflammatory markers were to be unchanged but differences in crying time**

observed, there are other potential explanations, for example effects on gastric emptying (Indrio et al, PMID 188382). Will they consider that the tested biomarkers performed are enough to assess intestinal inflammation? **To our knowledge, there is no noninvasive marker to date that is more sensitive than fecal calprotectin.** Will authors consider to investigate the intestinal mucosal microenvironment in infants with colic? **The intestinal mucosal microenvironment could be different from the fecal microbial communities, although there should be similarities. We are performing a larger study of the fecal microbiome in infants with colic. We would be interested in biopsy-generated information, which has been shown to be optimal for studies of inflammatory bowel disease, but our U.T. IRB (we suspect) would not likely approve endoscopy/colonoscopy for a self-limited condition in infants this young.**