Article details: 2016-0087	
Title	*Lactobacillus* probiotics in the prevention of *Clostridium difficile* associated diarrhoea: A systematic review and Bayesian hierarchical meta-analysis
Authors	Alison Sinclair MD, Xuanqian Xie MSc, Lama Saab MSc, Nandini Dendukuri PhD
Reviewer 1	Dr. Nathan Eveniew
Institution	McMaster University, Department of Surgery
General comments (author response in bold)	The authors present a systematic review and meta-analysis of 10 RCTs for probiotics in the prevention of Clostridium difficile-associated diarrhea (CDAD). They used primarily a Bayesian hierarchical approach for data synthesis with several sensitivity analyses and reported a pooled risk ratio of 0.25 (95% credible interval 0.08, 0.47; 95% prediction interval 0.02, 1.34), which suggests a significant but imprecise benefit. They conclude that there is considerable uncertainty regarding the efficacy estimate associated with lactobacillus probiotics and that the evidence is inconclusive and inadequate to support a policy concerning routine use of probiotics. Major comments: The authors limited this systematic review to only those probiotics containing Lactobacillus species, but the basis for this decision was not adequately described or addressed as a possible limitation. In fact, it seems unlikely that there would be sufficient biological, clinical, or methodological rationale to support this choice because a recent meta-analysis included studies of other probiotics, performed credible subgroup analyses according to pre-specified hypotheses (for example, S. boulardii vs. L. rhamnosus), and reported – based on evidence from twice as many trials as this review - that results were similar across subgroups (Johnston et al., 2012, Annals of Internal Medicine – reference #7). Therefore, the current study seems to be missing a large volume of relevant data, which could potentially influence both the point estimate (ie. accuracy) and the credibility/prediction intervals (ie. precision). **The exact biological mechanism by which probiotics work is in fact unknown (please see citations to Hickson et al and Allen et al in the manuscript). In our previous work (Dendukuri et al., CMAJ 2005, Dendukuri et al., Am J Gastroenterology, 2007), we have argued against pooling results across such a heterogenous group of studies as it is conceivable that the magnitude of the efficacy depends on the type of micro-organism (bacteria vs. fungus) or the age of the population (as children are naturally colonized with *C. difficile* bacteria unlike adults).** **The reviewer refers to a recent meta-analysis (Johnston et al, Annals of Int Med, 2013) studying the efficacy of probiotics for preventing *C. difficile* diarrhea that included studies of different types of probiotics and different age groups of patients, yielding a much larger number of studies than we have. Though the authors may not have found evidence of statistically significant differences in the pooled effect size between sub-groups, this does not imply that there are in fact no differences between them. Lack of statistical significance may simply be attributable to insufficient numbers of studies in each sub-group, small sample sizes of individual studies or the particular statistical approach used to model heterogeneity.** **Our objective was to estimate the efficacy of *lactobacillus* probiotics in adults. By including studies that strictly answer this question we increase the internal validity of our study. Clearly, our results are not generalizable to probiotics based on yeast or to pediatric populations. We also limited ourselves to studies published in peer reviewed journals to minimize concerns of risk of bias and publication bias. We say so in the Discussion section.** The authors stated that the evidence is "inconclusive and inadequate", but they have not implemented a system to formally and transparently

	rate confidence in the pooled effect estimates. Confidence ratings are important because they inform evidence users about the quality of the evidence available for clinical decision-making, integrating factors such as study design, risk of bias, imprecision, inconsistency, indirectness, and publication bias. The Grades of Recommendation, Assessment, Development, and Evaluation (GRADE) approach has been adopted by more than 70 major health research organizations for this purpose, including the Cochrane Collaboration, the World Health Organization, and the American College of Physicians; several meta-analyses recently published in CMAJ have also used the GRADE approach. Without use of GRADE (or another similar system, but GRADE is preferred), their conclusions are not supported by their results. **We have now implemented the GRADE approach to rate the quality of the evidence regarding the efficacy. In addition to what the GRADE guidelines typically require, we have added: i) the prediction interval, ii) the range of the risk of CDAD in the control group, so as to explain our concerns about inconsistency.** The authors chose not to pool data for harms or adverse effects, stating that previous meta-analyses have already done this, but this decision contradicts their rationale for updating their meta-analysis on the basis of new data being available. Given that the recent trial (Allen et al., 2013, Lancet) is so much larger than the other trials, the pooled effect estimates for harms or adverse effects should also be updated so that decision-makers can accurately consider the balance of desirable and undesirable effects of alternative management strategies when making individual clinical decisions or preparing guidelines. **Our focus was on pooling the efficacy estimates using appropriate methods and to use the pooled estimates to make inferences. We now make this clear in the objectives and Discussion of the article and drop any mention of pooling adverse events.** The authors stated that there is "considerable uncertainty" in their pooled effect estimates due to heterogeneity, but the magnitude of heterogeneity according to their Bayesian approach hasn't been clearly presented. Thus, although the authors used meta-regression to explore potential sources of heterogeneity, it is unclear whether there was an actual need to do so. According to their conventional metric, which is likely to be more familiar to most readers, it actually appears that there was no significant heterogeneity (Q-test p-value was not significant; I-squared percentage not presented), which suggests that conclusions do not match their results. Although the authors state the conventional metrics are at risk of underestimating heterogeneity, they are nonetheless widely accepted and this p-value was not borderline. **We now present the Bayesian approach and the conventional DerSimonian-Laird approach side-by-side in Table 3 to make the comparison between the two approaches more clear. The Q-test is well known to have low power (see the Cochrane Collaborations handbook for Systematic Reviews of Interventions). Therefore, that the Q-test is not statistically significant does not necessarily mean heterogeneity is absent. It may simply mean that we lack enough studies to detect it. That seems to be the case in the meta-analysis at hand. We agree, the conventional DerSimonian-Laird model is widely accepted. However, it is not necessarily the best model, and this has been recognized repeatedly in the statistical literature and more recently in clinical journals as well (please see citation to Cornell). This is what we would like to draw attention to.** The authors have not presented the pooled effect estimates in absolute measures (ie. absolute risk and/or numbers needed to treat). Relative measures of association are often challenging to interpret and in some instances may be misleading (for example – relevant to this study - high RR in presence of very low baseline risk can yield unimportant ARR/high NNT). Given that the authors state that the effect of probiotics might vary depending on different populations' baseline risks for CDAD, it

	would be ideal to present different estimates of absolute effects for each set of population characteristics. **We have now added meta-analyses of the risk difference within low and high prevalence groups.** The Bayesian approach to pooling data is well described but certainly not conventional and perhaps even controversial, and many readers are likely to have difficulty interpreting the presentation of the results due to unfamiliarity. Further justification for this particular choice and a balanced discussion of potential advantages and disadvantages in terms that readers will understand is required, particularly given that their sensitivity analysis with the conventional DerSimmonian-Laird approach found a somewhat greater RR (0.42) with narrower confidence intervals (0.29 to 0.59). **We have added more details on the Bayesian approach – general details appear in an information box and specific results for the meta-analysis at hand are juxtaposed with the results based on the DerSimonian-Laird method in Table 3. We have also added to the Discussion references to articles that have appeared over the last two decades by respected scientists arguing in favour of Bayesian methods for both medical research and health technology assessment.** Minor comments: Risk of bias assessments with each individual domain of the Cochrane Risk of Bias tool should be presented separately for each study, rather than aggregate A/B/C. **This has now been done.** AAD does not seem to be defined in the main text. **This has now been corrected.** Table 1 – Unclear how or why three of the studies have zero follow-up. **This has been corrected.** "Double-blinding" – this terminology is confusing because there are several groups that could be blinded in RCTS (clinicians, patients, outcome assessors, data collectors, data analysts, manuscript writers, etc). The authors have not reported whether any trials were excluded because they were unblinded and/or "blinded" but not "double-blinded". If so, it would be preferable and conventional methodology to include all eligible RCTs regardless of blinding, then report complete risk of bias assessments, then explore potential heterogeneity according to a pre-specified subgroup hypothesis given that the authors believe blinding to be an important issue, then rate confidence in the pooled effects estimates accordingly using the GRADE system. **We now make clear that by double-blinding we refer to blinding of the patient and the outcome assessor, and that the reason we consider this important is because it allows us to assume that any missing *C. difficile* test results are missing completely at random. Though the GRADE approach advocates pooling all studies first and downgrading later, we prefer to take the more conservative approach of excluding from our primary analysis those studies which are known at the outset to have a possibility of bias. This is similar to our reasoning to limit the inclusion criteria to published studies of *lactobacillus* probiotics in adults.** **As the reviewer acknowledges in an earlier comment, the GRADE approach is but one approach for summarizing evidence, which has been widely publicized. It may be a very helpful starting point, listing some of the key issues to be considered, and it offers the important advantage of transparency in reporting and reasoning. None the less, like all**

	guidelines and checklists, it does not provide a substitute for critical thinking. Others have commented (Weed, 2013, ACP Journal Club) that perhaps the trial by Allen et al. was negative simply because AAD and CDAD were much less frequent than expected. This issue warrants further discussion because it is plausible and it opposes the authors' fundamental assertion that data from this trial contradict previous meta-analyses. **As the reviewer has noted in a previous comment, we did investigate the hypothesis that the prevalence of CDAD is associated with the risk ratio. This was done descriptively via a L'Abbé plot and a meta-regression analysis. Our investigation of the relation between the prevalence of CDAD and the risk ratio was indeed an attempt to understand why the conclusions of the Allen trial and the earlier meta-analyses differ. None the less, it remains a fact that the Allen trial reached a negative conclusion whereas some earlier meta-analyses reached a positive conclusion favoring probiotics.**
Reviewer 2	Dr. Paul Arora
Institution	Public Health Agency of Canada, Office of Biotechnology, Genomics and Population Health
General comments (author response in bold)	Abstract Specify that the results of a the more recent trial found no benefit of lactobacillus probiotics on the SAME outcome as the recent meta-analyses have been studying, namely, CDAD. **This has been added.** Under methods: "We carried out an updated systematic review and meta-analysis of…[insert exposure/outcome relationship]...using a ..." The outcomes of interest are not clear in the abstract. **The outcome has been clarified in the abstract.** Manuscript body Page 6 line 3: 1) What was the effect on your results of including active treatment at "any dose"? **If in fact there was a variation in the efficacy according to the dose of the treatment, then our pooled risk ratio would be an aggregate of more potent and less potent doses.** 2) You have conducted meta-regression using a categorical (low/high) cutpoint for dose. Why not use a continuous dose variable to measure the effect of dose on your effect estimates more precisely? **We implemented the suggested model with a continuous measure of dose and found no apparent relation as in the case of the dichotomous variable. The regression coefficient was 0.002 (-0.03, 0.05) suggesting no apparent association between dose and the log risk ratio. We had to drop one outlying study (Selinger) in this analysis, which reported 10 times the dose of other studies, to avoid model convergence problems.** Page 6 line 8: "...defined as diarrhea…" What specific definitions did you apply for diarrhea or did you take ANY definition of diarrhea? That is, what definition of diarrhea would you have excluded? It seems you included ANY definition of diarrhea. The last two studies included in the review did not list definitions for diarrhea. Did you conduct sensitivity analysis to assess the effect of removing the studies without a specific definition of diarrhea (defined by you or some generally agreed upon standard)? **We have clarified that we did indeed include studies even if they did not provide a definition of diarrhea. We have now added a sensitivity analysis limiting to those 5 studies that provided a more strict definition of diarrhea of ">= 3 liquid stools". Our inferences were**

	unaltered. Page 6 line 35: AAD has not been defined in the text yet. Define in full before using the short form. I assume "Antibiotic-associated diarrhea". **We have now defined AAD as Antibiotic-associated diarrhea.** Page 6 line 49 (comment applies to whole paragraph): I believe this paper hinges critically on the reader understanding Bayesian approach; I would imagine the average CMAJ reader does not. The manuscript makes clear the qualitative difference in results depending on applying Bayesian methods. The authors should define what priors are and what a "vague" prior is. I would also recommend a box where a paragraph can be dedicated to explaining what the Bayesian statistical approach is. This may seem like too much information but I interpret this manuscript as demonstrating the different conclusions that are reached when the Bayesian approach is used; the methods used are not made clear enough given their importance to the results and interpretation. **We have now added a box explaining the Bayesian approach and illustrating the influence of the prior distribution in various prototypical situations.** Page 7 line 7: (comment applies to whole paragraph): Credibility analysis seems like a novel tool to apply here and the authors have done a good job at explaining a difficult method. I would like to see some more explanation of the "critical sceptical prior distribution". I would strongly encourage the inclusion of a graphical method to communicate how this procedure works and its results interpreted. **We have also added several prototypical plots within the box on Bayesian analysis that will help the reader to understand the role of the prior in a Bayesian analysis. These plots would also assist the reader in understanding the reasoning behind the credibility analysis.** Page 10 line 23: While it would be premature to recommend it in "all settings", perhaps lactobacillus probiotic treatments may be effective in settings in high burden of disease/force of infection? The authors have not ruled this out. **Our overall conclusion is that the available evidence is inconclusive. We believe an important reason for this is that the number of CDAD cases across the studies is very small. We seek to draw attention to this point by way our credibility analysis.** **We feel that it is difficult to make any recommendation even for high burden settings based on the available evidence. One study (by Gao et al.), which had the greatest weight in the meta-analysis, was conducted in a high burden setting and had a positive result despite the fact that it had a relatively small sample size. Though it had a high rating for the quality of the evidence, there is concern for the generalizability of these results as it is the only study with a very high risk of CDAD in the placebo group and it was supported by the manufacturer.** Page 11 line 14: The authors state the included trials contain some important exclusions of patients at high risk of infection. Did this reduce the effect estimate of the treatment as the authors suggest that the treatment may be effective in high burden of infection settings? Can the authors highlight which studies excluded high risk patients and conduct analyses on those that kept high risk patients in? If ALL studies excluded high risk patients then the authors should comment on this potential dampening on of the effect estimate and what direction the effect estimate may move to if they were included.

	We had included this sentence previously when referring to the risk of adverse events. By high risk of infection we meant infections other than *C. difficile* diarrhea. Since we no longer report on adverse events we have dropped this sentence. Table 2: Spell out AAD and CDAD in the footnotes. **We have now expanded the abbreviations AAD and CDAD in the footnotes.** Page 20 FLow Diagram: The first two boxes add up to 17 not 15. I assume there were 2 duplicates? This should be indicated somehow. **The PRISMA diagram has been redone at the request of the 3rd reviewer.**
Reviewer 3	Dr. Paul Ronksley
Institution	University of Calgary, Community Health Sciences
General comments (author response in bold)	Comments to the Author The purpose of this systematic review and meta-analysis was to address the association between Lactobacillus species probiotics in the prevention of Clostridium difficile associated diarrhoea. More specifically, the potential influence of between-study heterogeneity on the overall conclusions drawn from previous meta-analyses using standard random effects techniques. In this manuscript, the authors use Bayesian methods to assess this association. They identified 10 studies that report on this association and conclude that there is considerable uncertainty regarding the efficacy estimate associated with lactobacillus probiotics due to high heterogeneity between studies. This is a very interesting and highly relevant topic with respect to whether policies should support the routine use of probiotics among hospitalized patients. The methods used in this manuscript are extensive and complex. I appreciate the use of this approach as a step-forward in representing statistical uncertainty. I do have some concerns around the methodology and overall reporting/interpretation of the findings. Introduction 1. Authors indicate that they have updated their previous systematic review of probiotics in the prevention of CDAD and conducted a meta-analysis of Lactobacillus species probiotics, applying Bayesian hierarchical methods. I would argue that this manuscript is not an update of their previous work since a new methodology is used here, a single organism is considered, and the focus here is on prevention of CDAD not treatment. The authors should consider the full PRISMA guidelines as opposed to those for an update. **We agree with the reviewer and have now followed the full PRISMA guidelines in preparing Figure 1.** 2. The objective of the study needs to be clarified in the introduction and reflected in the abstract as well. It is not clear whether the purpose is to conduct a meta-analysis of probiotics for the prevention of CDAD incorporating the new study by Allen et al. evaluating lactobacillus acidophilus using Bayesian methods OR whether the purpose is to evaluate between-study heterogeneity using Bayesian methods and using probiotics for the prevention of CDAD as the clinical example. It appears to be the former (but the background section from the abstract would make you think otherwise). **The reviewer is correct that our primary objective is to carry out a meta-analysis of probiotics for preventing CDAD. A secondary objective is to pay particular attention to improving the estimation of between-study heterogeneity than has been done previously. We have tried to emphasize these two separate goals with greater clarity in both the main text and abstract.** Methods 3. Paragraph 1. Although authors indicate that they extended the search from their previous study, the search for this manuscript appears

different from the original study as it is currently written. References 14 and 15 are not published or available to the public and therefore cannot be reviewed for comparison of the search strategy. Therefore, I suggest the authors re-write this paragraph to describe the differences between their original study and the current study in terms of their search strategy. See subsequent comments.

a. Paragraph 1. Some of the online databases overlapped with the original study (e.g. PubMed, EMBASE) while other databases were added (e.g. AMED, CINAHL). Given these differences, it may be more appropriate to complete a full PRISMA flow diagram and checklist rather than an adapted one.

We agree with the reviewer. Though we reference our previous work (which is available to the public at http://www.mcgill.ca/tau), we treat the current review as resulting from a new literature search and accordingly have prepared the full PRISMA flow diagram (Figure 1).

b. Authors indicate that their search strategy is provided in the Supplementary Material; however this was not included for review. Please ensure that their complete search strategy is provided (i.e. the original and any updates) for at least one database (e.g. PubMed or EMBASE).

We have now corrected this and added the strategy in PUBMED to the Supplementary Material.

4. Authors should follow the PICOD/PICOS when describing their inclusion criteria in the methods section. The authors did not explicitly state what the control or comparator group was in the included studies (though it appears to be placebo). The primary outcome defined in Line 34, paragraph 1 of statistical methods should be listed with inclusion criteria and definition of CDAD (lines 4-10, paragraph 2 of Methods). It is not clear from the listed inclusion criteria that the authors were looking for nosocomial CDAD. This should be explicitly stated as part of the primary outcome. Definition of CDAD should also include definition for considering it nosocomial.

We now use the PICOS format when stating our Objective as well as in the Methods section. The outcome is indeed nosocomial CDAD.

5. The authors assessed the risk of four biases described by the Cochrane Collaboration for each study and then classified the studies into three groups based on the number of sources of bias. This type of grouping is fairly arbitrary and may not be as informative as a table that allows the reader to assess the quality of each individual study. This could be added as a supplementary table and would provide more information about the potential biases present and how they may contribute to the heterogeneity.

We have now added the detailed presentation of the risk of bias in individual studies to the supplementary material (Supplementary Table 3).

6. AAD is mentioned in Statistical Methods for the first time and it is not clear to the reader how it relates to inclusion criteria. Please define AAD at its first occurrence (Paragraph 3).

AAD is now defined at the first occurrence.

7. L'Abbe plot was used to examine the relationship between the risk of CDAD in the placebo group of included studies. While this is a great approach – it may not be familiar to many readers and therefore a brief description of its function may be helpful for the interpretation of Figure 2 (of course space pending).

We have added more description of the L'Abbé plot.

	8. For the credibility analysis, where is their explicit statement of their prior knowledge? Is it the 0.6-1.7 that is listed in the results section? The methods section should indicate what the prior knowledge or distribution they are considering and the results should indicate what the posterior interval is given the prior interval and the data. **The primary analysis used vague prior distributions on the unknown parameters (the pooled risk ratio and the between-study variance) as explained in the Methods section. The corresponding posterior credible intervals appear in the Results section.** **In the case of the credibility analysis, the critical sceptical prior distribution is in fact a result. Accordingly it is presented in the Results. The critical sceptical prior is determined by asking "What prior distribution centred at RR=1 would need to be combined with the observed results to yield a posterior credible interval for the risk ratio that includes 1?" For our problem it turned out to be a prior distribution whose 2.5% and 97.5% quantiles are 0.6 and 1.7, respectively. We have now tried to explain this better and distinguish our main results from the credibility analysis.** 9. The authors conducted a meta-regression evaluating the effect of study quality, prevalence of CDAD, composition of probiotic, dose of probiotic and industry funding on the pooled risk ratio. Have the authors considered evaluating length of follow-up and treatment as other potential sources of heterogeneity? Of course – the fact that meta-regression is underpowered will likely show that these variables are not significant predictors – but it might be interesting to explore. **We did not adjust for the two covariates mentioned by the reviewer as it was not possible to quantify them in a systematic way across studies. For length of treatment duration, many studies reported it as "Antibiotic + x days". So the duration of treatment would have varied for each patient depending on the number of days for which the antibiotic was prescribed. The length of follow-up, was variously defined with respect to the baseline, to the end of antibiotic treatment or to the end of probiotic treatment. We now mention this as a limitation in the Discussion.** 10. Overall, the methods section could be improved with the inclusion of sub-headings. For example: Search Strategy, Inclusion Criteria, Data Abstraction, Evaluation of Bias, Statistical Methods, Credibility Analysis, Sources of Heterogeneity, Sensitivity Analysis. The incorporation of descriptions for each type of analysis (particularly the credibility analysis and Bayesian modeling) is informative. I don't know if it's possible to make this language more attainable to a lay audience – probably not given the nature of the analysis. **Sub-headings have been added to the Methods and Results section for greater clarity.** 11. Paragraph 3, Line 27. "we used a funnel plot…"- This should be moved down under Statistical Methods. **We have done as requested.** Results 12. Paragraph 2. The date range of the included studies could be included after the first sentence. **The date range has been added.** 13. Please provide the acronym for 'colony forming units' (i.e. CFU) since it used in other parts of the paper. **We have added the expansion of CFU when it is used first.** 14. How did they come to the critical sceptical prior range of 0.6-1.7?

	See comment 8 above. Further details about this range are required. **We have added more details to explain the derivation of the critical sceptical prior.** Interpretation 15. Given that this is a relatively new approach to conducting meta-analyses, the authors should discuss the limitations of Bayesian methods and the credibility analysis and how they justify their use of each and the prior distributions used at some point in the discussion. **We have added a box summarizing key points on the Bayesian approach in response to Reviewer 2.** References 16. References 8 and 11 repeated **This has been corrected.** Table 1. Please clarify what is meant by "Elderly" under "Sample Size Pro; Pla" for the Plummer 2004 study. **This was a mistake which has now been corrected by replacing the word Elderly with the sample size in probiotics and placebo groups in that study.** Table 3. For the results of the meta-regression, it might be more informative to see the stratum specific estimates with the actual relative risk and its credible interval instead of the regression coefficient so that the results in this table are comparable to the meta-analysis results. **We have changed the presentation of the meta-regression as requested (please see revised Table 4).**