PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Vitamin D supplementation to improve pregnancy and perinatal outcomes: an overview of 42 systematic reviews
AUTHORS	Bialy, Liza; Fenton, Tanis; Shulhan-Kilroy, Jocelyn; Johnson, David; McNeil, Deborah A.; Hartling, Lisa

VERSION 1 – REVIEW

REVIEWER	Kathryn Kaiser University of Alabama at Birmingham
REVIEW RETURNED	18-Jul-2019

GENERAL COMMENTS	This overview of systematic reviews is reporting the results of the examination of available reviews addressing observational or RCT evidence of the effect of Vitamin D supplementation or status on pregnancy/perinatal outcomes. The methods appear to be generally well-described and of sufficient rigor. Major and minor comments are grouped below in no particular rank order, and when applicable, page:line numbers are referenced at the beginning of the comment. Major concerns/suggestions: 1) The title, while reflective of the authors' perspective given the findings, is not informative for readers at a glance. A more accurate and objective title that reflects the state of the evidence would be more informative, e.g. "No high quality evidence supports Vitamin D supplementation to improve pregnancy/perinatal outcomes: an overview of 42 systematic reviews". 2) It is not clear from references to assessment of publication bias whether the authors applied any objective measure to reviews, e.g. the ill-advised reliance on funnel plots or the better use of Begg's or Egger's tests. One question the authors could settle would be a comprehensive test across all included RCTs. Without this, the comments regarding publication bias are not well-supported. See page 20:31 – perhaps further insight from this paper will illuminate: https://onlinelibrary.wiley.com/doi/abs/10.1002/sim.698 Minor concerns/suggestions: 1) No PRISMA checklist was provided. 2) The abstract objective statement should be clarified with more precision and this should be echoed more clearly in the introduction. 3) The authors opted to use AMSTAR instead of AMSTAR 2 - https://www.bmj.com/content/358/bmj.j4008 If this was considered and not chosen for some reason, it should be explained. Perhaps the use of AMSTAR 2 on the higher quality reviews would add helpful information? 4) The last paragraph of the introduction was difficult to follow in a linear flow – perhaps a final sentence stating something like "We aimed to address this state of the literature… by doing … in order to summarize the quantity and levels of evidence to inform future

	research directions." 5) 9:5-8 – the statement about conversion to risk ratios and the random effects model implies something that does not later appear in the results clearly. Please elaborate. 6) 19:47 – This sentence needs to be clarified/rephrased – is it a commentary about heterogeneity among the RCT designs and purposes that should not be taken as a unified body of evidence? 7) Supplementary Table 3 – suggest changing "ca" for can't answer to something else – either the authors report they did something (y), they explicitly state they did not do something (n) or it was not reported (NR).

REVIEWER	Thalia Sparling Tufts University, Friedman School of Nutrition Science and Policy, USA
REVIEW RETURNED	01-Sep-2019

GENERAL COMMENTS	Overall, the research objectives are clear, the methods are sufficiently presented and the paper is well organized into outcomes and study designs. My biggest qualm is that the authors have based all their assessment of results on the subjective assertion that risk ratios between 0.5 and 2 are not large. I not agree with this statement as an absolute and it seems to be a problem of ecological fallacy and a misunderstanding of epidemiology. Whether risk ratios are large or small depends on the population attributable risk fraction (based on the prevalence of the outcome in the population) and the severity of that outcome. Furthermore, I assume the overarching practical goal of this paper is to assess/consider perinatal supplementation schemes. The importance of a 10% absolute or even relative change in these outcomes is not necessarily insignificant. I recommend reconsidering the subjective 'cut-offs' and remove the interpretations throughout the paper based on those cut-offs unless the authors can reasonably justify these choices from a population health perspective. I appreciate the conservative interpretation of the evidence, but in this rare case, I think it might be too conservative based on the authors assumptions. Those assumptions would need some evidence to back them up or be taken out. Other small comments are attached in the PDF proof, but a few are: - It might be good to explain a bit more clearly that one SR can have different GRADE rankings based on different outcomes and give an example. - there are a few typos -- see manuscript comments (typo in study selection of the abstract - The authors mention publication bias a lot. It is a real thing, but the suggestion/possibility of publication bias is different than evidence of it through a forest plot, etc. In the paper, publication bias may be over-emphasized without justification that these claims are from actual evidence presented by the SRs. If they all showed evidence of publication bias, that is one thing. If they said 'it's possible' and that's a limitation, I don't think we should discount existing evidence too much at this point in this topic. - Some of the authors' statements, including their general assertion that RCTs will solve our problems and that observational studies are not believable (Sometimes true but not always), are a bit strong. I have made some specific comments in the manuscript – please contact publisher for this file, but it might be worth being a bit more careful with the language.

REVIEWER	Calvin J. Hobel, MD Cedars Sinai Medical Center Los Angeles, California USA
REVIEW RETURNED	12-Sep-2019

GENERAL COMMENTS	Background Comment: I do not think the authors understand how pregnancy affects the outcome of pregnancy. First, the authors mention on page 10 (line 49) with the statement “generally the populations studies were healthy without preexisting conditions”. Approximately 30% of women have asymptomatic preexisting conditions (asymptomatic) making them high risk when they get pregnant. The authors should read the paper by (Williams D. Pregnancy a S.tress Test for Life. Current Opinions Obstet. Gynecol. 2003 15:465-471) which means that there is no other time in the life of a woman where the immune system changes in early pregnancy to prevent rejection of the fetus and where there are tremendous changes in the cardiovascular system to assure adequate blood flow to prevent poor fetal grow (low birth wt. / intrauterine growth restriction). QUESTION #3 #6, The issue of perinatal outcomes: Poor pregnancy outcomes occur in about 25-30% of pregnancies. The authors initially lists 6 poor pregnancy outcomes (preterm birth, preeclampsia, gestational diabetes, SGA, Low Birth wt. (<2500grams) still birth & C-section. One of the most important problems in poor pregnancy outcome is the cause of poor fetal growth which is a common finding in poor pregnancy outcome except for gestational diabetes. I THINK THE AUTHORS SHOULD HAVE DEFINED THE SUBJECTS RISK FACTORS ASSOCIATED WITH RISK of poor pregnancy outcome, SUCH AS OBESITY, BMI & RACIAL STATUS (IT IS NOW WELL KNOWN THAT DARK SKINNED WOMEN (AFRICAN AMERICAN & HISPANICS ARE MORE LIEKLY TO BE VITAMIN D DEFICIENT. COMBING THE SGA & Low Birth Wt. into one category of POOR FETAL GROWTH, MAY HAVE IMPROVED THE RESULTS OF THE TWO GROUPS (TABLE 4 PAGE 38 RCT’S) & OBSERVATIONAL STUDIES PAGE 41. QUESTION #8: Are the references up to date? I will list several papers the authors may fine interesting and helpful in understanding the importance of vitamin D: #1) Accortt, EE, Miroka J, Dunkel Schetter & Hobel CJ. Adverse Perinatal Outcome and Postpartum Multisystem Dysregulation. Adding Vitamin D Deficiency to Allostatic Load. Maternal Child Health Jour. Vol 17 #7m 2017. This paper documents that adding vitamin D deficiency as an additional metric to the classic allostatic load score used to look at metabolic risk of cardiovascular disease. #2) Liu PT, Stenger S, Li Huiying et al. Toll-Like Receptor Triggering of Vitamin D Mediated Antimicrobial Response. Science Marsh 24, vol 311, 2006. This paper identified for the first-time circulating vitamin D initiates the production of cathelicidin the natural antibiotic. #3) Tanz LJ, Stuart J., Williams Pl, et al. Preterm delivery & Maternal Cardiovascular Disease in Joung Women & Middle Age Adult Women. Circulation. 2017; 135:578-589. This paper links having a preterm birth significantly increases early the risk of cardiovascular disease in women. #4) Berg AH, Powe CE, Evans MK, et al. 24-25 -dihydroxyvitamin D3 and Vitamin D Status of community Dwelling Black and White Americans. Clinical Chem. 2015:61(6):877-884. This paper begins to look at a metabolite of vitamin 25-D to assess a new biomarker using the ratio of 24,25 D to 25 D (24,25D/25D to be used to assess

	vitamin D status in patients of color. QUESTION #9; Do the results address the research question and results? #1) The results are primarily epidemiologic. There is no attempt to consider mechanisms for why vitamin D is important. There are many issues today about the association between obesity whereby vitamin D is soluble in fat. In addition; obesity and diabetes are associated with a great risk of inflammation and cardiovascular disease. Systems Biology is becoming import to understand diseases causing poor pregnancy outcome, but the cause maybe early cardiovascular disease in women which is asymptomatic until pregnancy is the test to assess risk of poor pregnancy outcome. #2) There is an issue about the amount of vitamin D to be used as a supplement or to treat deficiency (if present). The authors mention on page 10 that the dose ranged from 200 units the 5000 units daily or 714 to 50,000 IUs and very high doses. It is well known that if a subject is deficient it takes at least 3000 IU daily to bring levels up to normal over a period of time. THUS, AS FAR AS STUDY DESIGN ,STUDIES SHOULD START WITH AT LEAST 3000 IU PER DAY (no less). IF STUDIES Do NOT GIVE SUFICIENT DOSES THEY SHOULD NOT BE INCLUDED IN STUDIES QUESTION #10: Are the research questions and results presented clearly? #1) To meet the design of their study and the poor design of most vitamin D studies being reviewed, the authors present the data they have. COMMENTS ON CONDLUSIONS: #1) On page 21 (line 35) the authors state that credibility of the evidence in this field is compromised by the potential for publication and reporting biases (at this point I think the authors should add "and poor study design"). #2) On the same page (line10-11) the authors mention the issue of confounding variables. I am concerned about not looking at risk factors present in most of the poor outcomes mentioned by the authors some of which are confounding variables. Some of these risk factors are present in many of the poor outcome reported and are part of the disease process. I think the issue is related to systems biology and the mechanism of theses disease. I believe and supported by good studies that vitamin D is important for the brain, heart, gastrointestinal tract, muscles, bone and the immune system (the latter to reduce infections). I listed an important paper on how vitamin D is important for the production of cathelicidin in macrophages.

VERSION 1 – AUTHOR RESPONSE

Reviewer 1

Comment	Reply
The title, while reflective of the authors' perspective given the findings, is not informative for readers at a glance. A more accurate and objective title that reflects the state of the evidence would be more informative, e.g. "No high quality evidence supports Vitamin D supplementation to improve pregnancy/perinatal outcomes: an overview of 42 systematic reviews".	This is an excellent suggestion which we have implemented, thank you.
It is not clear from references to assessment of publication bias whether the authors applied any objective measure to	We included the data that was presented in each of the SRs; therefore,

reviews, e.g. the ill-advised reliance on funnel plots or the better use of Begg's or Egger's tests. One question the authors could settle would be a comprehensive test across all included RCTs. Without this, the comments regarding publication bias are not well-supported. See page 20:31 – perhaps further insight from this paper will illuminate: https://onlinelibrary.wiley.com/doi/abs/10.1002/sim.698	we relied also on whether and how each SR assessed for publication bias. In many cases, there were too few studies for any given outcome in a review to assess for publication bias (either through funnel plots or statistical tests). In any case, we have removed references to publication bias in the text of the results. We are concerned, however, that there is an issue of reporting bias; we have described this in the last paragraph of the discussion. Specifically, for many outcomes SRs were only able to identify and pool data for a small number of studies, despite the large number of studies overall in this area.
No PRISMA checklist was provided.	PRISMA checklist has been included as a supplementary file.
The abstract objective statement should be clarified with more precision and this should be echoed more clearly in the introduction.	The objective in the abstract and the introduction have been modified.
The authors opted to use AMSTAR instead of AMSTAR 2 - https://www.bmj.com/content/358/bmj.j4008 If this was considered and not chosen for some reason, it should be explained. Perhaps the use of AMSTAR 2 on the higher quality reviews would add helpful information?	We used AMSTAR as AMSTAR 2 had not yet been published when we began this study. We are not convinced that AMSTAR 2 assessments for the higher quality reviews would provide additional useful information. We feel that application of GRADE highlights the strengths and limitations across this body of evidence.
The last paragraph of the introduction was difficult to follow in a linear flow – perhaps a final sentence stating something like "We aimed to address this state of the literature… by doing … in order to summarize the quantity and levels of evidence to inform future research directions."	Good suggestion, thank you. We have revised the last paragraph of the introduction.
9:5-8 – the statement about conversion to risk ratios and the random effects model implies something that does not later appear in the results clearly. Please elaborate.	Footnotes in Table 1 and 2 have been revised to clarify those that were not converted to risk ratios. A sentence has been added to results (page 11).
19:47 – This sentence needs to be clarified/rephrased – is it a commentary about heterogeneity among the RCT designs and purposes that should not be taken as a unified body of evidence? "The evidence contributing to the existing SRs varied widely in design and purpose."	We have expanded on this sentence to clarify. Also the remainder of the paragraph expands on this initial statement to explain it more fully.
Supplementary Table 3 – suggest changing "ca" for can't answer to something else – either the authors report they did something (y), they explicitly state they did not do something (n) or it was not reported (NR).	The AMSTAR guidelines provide 'Can't Answer' as the option in their checklist if 'Yes' or 'No' cannot be indicated. Shea BJ, Hamel C, Wells GA, Bouter LM, Kristjansson E, Grimshaw J, Henry DA, Boers M. **AMSTAR is a reliable and valid measurement tool to assess the methodological quality of systematic reviews.** J Clin Epidemiol. 2009 Oct; 62(10):1013-20. PMID: 19230606

Reviewer 2

Comment	Reply
My biggest qualm is that the authors have based all their assessment of results on the subjective assertion that risk ratios between 0.5 and 2 are not large. I not agree with this statement as an absolute and it seems to be a problem of ecological fallacy and a misunderstanding of epidemiology. Whether risk ratios are large or small depends on the population attributable risk fraction (based on the prevalence of the outcome in the population) and the severity of that outcome. Furthermore, I assume the overarching practical goal of this paper is to assess/consider perinatal supplementation schemes. The importance of a 10% absolute or even relative change in these outcomes is not necessarily insignificant. I recommend reconsidering the subjective 'cut-offs' and remove the interpretations throughout the paper based on those cut-offs unless the authors can reasonably justify these choices from a population health perspective. I appreciate the conservative interpretation of the evidence, but in this rare case, I think it might be too conservative based on the authors assumptions. Those assumptions would need some evidence to back them up or be taken out.	We have removed comments in the results section regarding magnitude of effect.
It might be good to explain a bit more clearly that one SR can have different GRADE rankings based on different outcomes and give an example.	In the methods section on page 9 a sentence has been added to describe the GRADE process.
there are a few typos -- see manuscript comments (typo in study selection of the abstract	Missing word in study selection has been included, and all other comments in manuscript have been fixed.
The authors mention publication bias a lot. It is a real thing, but the suggestion/possibility of publication bias is different than evidence of it through a forest plot, etc. In the paper, publication bias may be over-emphasized without justification that these claims are from actual evidence presented by the SRs. If they all showed evidence of publication bias, that is one thing. If they said 'it's possible' and that's a limitation, I don't think we should discount existing evidence too much at this point in this topic.	We have removed comments about publication bias in the results section.
Some of the authors' statements, including their general assertion that RCTs will solve our problems and that observational studies are not believable (Sometimes true but not always), are a bit strong. I have made some specific comments in the manuscript, but it might be worth being a bit more careful with the language.	We have reviewed the text to ensure that we have not made any strong claims such as “RCTs will solve our problems” and “observational studies are not believable”. Our premise is that the designs help answer different questions: RCTs are more reliable to answer questions of effectiveness and observational studies can address questions of association (but not often causation).

Reviewer 2 - Comments in PDF

Comment	Reply
Page 4; line 12: “low certainty evidence” seems	“Low quality studies” is different than “low

like jargon and is also repetitive, consider revising to “low quality studies”	certainty evidence” – certainty is a concept promoted by the GRADE working group to reflect the quality of the evidence which includes quality of studies (risk of bias/study limitations) as well as other factors including imprecision, indirectness, inconsistency, and publication/reporting biases. In any case we have changed “low certainty evidence” to “low quality evidence”.
Page 9; line 40: most epidemiologists will disagree that, across the board, a risk ratio of 0.5 to 2 is not large at a pop level. You should probably explain more….preeclampsia being half or twice as likely in a large sample is not a small effect. Neither pre-term birth or the others. This also depends on units of measurement – above or below dichotomous cut-off would mean a pretty big effect, no? Maybe I am misunderstanding you… Page 13, table 1 same comment on arbitrary cut-offs; this is not nothing unless you justify it based on literature Page 15; line 40: Again, I think clinical significance needs to be considered more in light of population attributable risk fractions and what this would mean on a population level, not an individual level since risk ratios are not applicable to individual patients anyway…….. Page 17; line 4 and 29: same comments Page 19; line 4-8: same comments	We have removed comments about the magnitude of effects.
Page 19; line 13 – 20: clunky sentence and could tone down this statement – one suggestion “leading to observed effects that were not always supported by RCTs which have a greater likelihood of establishing causal relationships”	Thank you for your suggestion. We agree, this sentence is too long and has been revised.
Page 21; line 8: change “controlled for” to “included”	This sentence has been revised.
Page 21; line 10: If they are decent RCTs, and if an RCT is an appropriate research tool for the research question, be careful not to wholesale the superiority of RCTs – not always true	Thank you – this sentence has been revised.
Page 21; line 19: I disagree with this statement – many important questions are not answerable by RCTs – you could soften this statement	Thank you – this sentence has been revised.
Page 22; line 50: what does “patient-important” mean - jargon	The wording has been revised.

Reviewer 3

Comment	**Reply**
I do not think the authors understand how pregnancy affects the outcome of pregnancy. First, the authors mention on page 10 (line 49) with the statement “generally the populations studies were healthy without preexisting conditions”. Approximately 30% of women have asymptomatic preexisting conditions (asymptomatic) making them high risk when they get pregnant. The	Thank you for your suggestion. We have relied on the inclusion criteria used for the individual studies included in the SRs.

authors should read the paper by (Williams D. Pregnancy a S.tress Test for Life. Current Opinions Obstet. Gynecol. 2003 15:465-471) which means that there is no other time in the life of a woman where the immune system changes in early pregnancy to prevent rejection of the fetus and where there are tremendous changes in the cardiovascular system to assure adequate blood flow to prevent poor fetal grow (low birth wt. / intrauterine growth restriction).	
The issue of perinatal outcomes: Poor pregnancy outcomes occur in about 25-30% of pregnancies. The authors initially lists 6 poor pregnancy outcomes (preterm birth, preeclampsia, gestational diabetes, SGA, Low Birth wt. (<2500grams) still birth & C-section. One of the most important problems in poor pregnancy outcome is the cause of poor fetal growth which is a common finding in poor pregnancy outcome except for gestational diabetes.	These 6 outcomes were defined a priori for this overview of reviews. However, based on this comment, we went back and reviewed all the outcomes in each of the systematic reviews and found that none of them included fetal growth as an outcome in their analysis. Therefore, it would not have been possible for us to include this outcome, even as a post-hoc analysis.
I THINK THE AUTHORS SHOULD HAVE DEFINED THE SUBJECTS RISK FACTORS ASSOCIATED WITH RISK of poor pregnancy outcome, SUCH AS OBESITY, BMI & RACIAL STATUS (IT IS NOW WELL KNOWN THAT DARK SKINNED WOMEN (AFRICAN AMERICAN & HISPANICS ARE MORE LIEKLY TO BE VITAMIN D DEFICIENT. COMBING THE SGA & Low Birth Wt. into one category of POOR FETAL GROWTH, MAY HAVE IMPROVED THE RESULTS OF THE TWO GROUPS (TABLE 4 PAGE 38 RCT'S) & OBSERVATIONAL STUDIES PAGE 41.	We were relying on analyses completed at the SR level which were based on analyses already completed by the primary studies.
Are the references up to date? I will list several papers the authors may fine interesting and helpful in understanding the importance of vitamin D: #1) Accortt, EE, Miroka J, Dunkel Schetter & Hobel CJ. Adverse Perinatal Outcome and Postpartum Multisystem Dysregulation. Adding Vitamin D Deficiency to Allostatic Load. Maternal Child Health Jour. Vol 17 #7m 2017. This paper documents that adding vitamin D deficiency as an additional metric to the classic allostatic load score used to look at metabolic risk of cardiovascular disease. #2) Liu PT, Stenger S, Li Huiying et al. Toll-Like Receptor Triggering of Vitamin D Mediated Antimicrobial Response. Science Marsh 24, vol 311, 2006. This paper identified for the first-time circulating vitamin D initiates the production of cathelicidin the natural antibiotic. #3) Tanz LJ, Stuart J., Williams Pl, et al. Preterm delivery & Maternal Cardiovascular Disease in Joung Women & Middle Age Adult Women. Circulation. 2017; 135:578-589 PubMed . This paper links having a preterm birth significantly increases early the risk of cardiovascular disease in women. #4) Berg AH, Powe CE, Evans MK, et al. 24-25 - dihydroxyvitamin D3 and Vitamin D Status of community Dwelling Black and White Americans. Clinical Chem. 2015:61(6):877 PubMed -884. This paper begins to look at a metabolite of vitamin 25-D to assess a new biomarker using the ratio of 24,25 D to 25 D (24,25D/25D to be used to assess vitamin D status in patients of color.	Thank you for these references; they are interesting reading. Our search was completed on January 1, 2019. Since our objective was to do a systematic review of systematic reviews and these studies are original cohort (Accort, Tanz), in vitro (Liu) and cross-sectional (Berg) studies, they would not have met our inclusion criteria.
The results are primarily epidemiologic. There is no attempt	You are correct that our study is not

to consider mechanisms for why vitamin D is important. There are many issues today about the association between obesity whereby vitamin D is soluble in fat. In addition; obesity and diabetes are associated with a great risk of inflammation and cardiovascular disease. Systems Biology is becoming import to understand diseases causing poor pregnancy outcome, but the cause maybe early cardiovascular disease in women which is asymptomatic until pregnancy is the test to assess risk of poor pregnancy outcome.	about potential mechanisms. We are trying to address the question of whether or not vitamin D is effective in improving pregnancy outcomes.
There is an issue about the amount of vitamin D to be used as a supplement or to treat deficiency (if present). The authors mention on page 10 that the dose ranged from 200 units the 5000 units daily or 714 to 50,000 IUs and very high doses. It is well known that if a subject is deficient it takes at least 3000 IU daily to bring levels up to normal over a period of time. THUS, AS FAR AS STUDY DESIGN, STUDIES SHOULD START WITH AT LEAST 3000 IU PER DAY (no less). IF STUDIES Do NOT GIVE SUFICIENT DOSES THEY SHOULD NOT BE INCLUDED IN STUDIES	Thank you for this comment as we are sure others will have these same questions. We added a sentence in methods at the end of analysis section. We have added in results based on dosing. Among the SRs of RCTs there were two SRs that conducted subgroup analyses based on dose: (1) high (>2000 IU/day) and low (≤ 2000 IU/day); (2) high (≥ 2000 IU/day) and low (< 2000 IU/day)
To meet the design of their study and the poor design of most vitamin D studies being reviewed, the authors present the data they have.	Yes, thank you
On page 21 (line 35) the authors state that credibility of the evidence in this field is compromised by the potential for publication and reporting biases (at this point I think the authors should add “and poor study design”).	We have removed this sentence based on another reviewer’s comment.
On the same page (line10-11) the authors mention the issue of confounding variables. I am concerned about not looking at risk factors present in most of the poor outcomes mentioned by the authors some of which are confounding variables. Some of these risk factors are present in many of the poor outcome reported and are part of the disease process. I think the issue is related to systems biology and the mechanism of theses disease. I believe and supported by good studies that vitamin D is important for the brain, heart, gastrointestinal tract, muscles, bone and the immune system (the latter to reduce infections). I listed an important paper on how vitamin D is important for the production of cathelicidin in macrophages.	We agree with you. If participants are not well randomized to intervention groups, there are likely to be differences in risk factors in the groups.

VERSION 2 – REVIEW

REVIEWER	Kathryn Kaiser University of Alabama at Birmingham, United States of America
REVIEW RETURNED	19-Dec-2019

GENERAL COMMENTS	No suggestions other than to check the median and range values reported for AMSTAR ratings (median 8, range 8-11). Mean instead of median?

REVIEWER	Calvin J. Hobel, MD Cedars Sinai Medical Center Los Angeles California USA
REVIEW RETURNED	12-Dec-2019

GENERAL COMMENTS	I think that the only significant finding of vitamin D deficiently in the SR of RCTs was Small for Gestational Age (<10th %tile) because this is the most serious poor outcome related to PTB (poor fetal growth). The study (Ref # 27) by Bi WG is real science. Also, what is interesting is that in the observational studies more significant findings were found which (I think) means that observational studies are done on different populations by different types of investigators. It was interesting that in the observation studies lower vitamin D levels were associated with SGA but NOT <2500 grams. Low Birth Weight (<2500grams is not a good index and percentiles should be used because percentiles by GA can give a spread (<10th) but also >10 to zero in on what is important between 10 and 25 as an example. The authors mention the “Acute Phase Reaction” several times. There are many examples of the acute phase reaction. One of the most important APRs is related to vitamin D. During an infection Toll receptor on a macrophage are stimulated by bacterial products which sends a signal into the nucleus to turn on a gene to make hydroxylase to add a OH ion to circulating vitamin D (providing there is no vitamin D deficiency) to produce active form of vitamin D [1-25 (OH)2 D] which goes back into the nucleus to turn on a second gene to make Cathelicidin that kills Tuberculosis Bacterium [that is why people went to Sanitoriums] to recover from TB. Today the issue of vitamin D studies is that there are two factors important as a cause of diseases. #1) It’s a combination of factors and the future of vitamin D studies to assess its importance will require a NEW APPROACH TO STUDIES which will require a systems biology approach. #2) Vitamin D and its role in health is undergoing dramatic change because of the issues of metabolism of vitamin D. Is it synthesis, metabolism, the role of binding protein or is it receptor genetics?

VERSION 2 – AUTHOR RESPONSE

REVIEWER #3 COMMENT: I think that the only significant finding of vitamin D deficiently in the SR of RCTs was Small for Gestational Age (<10th %tile) because this is the most serious poor outcome related to PTB (poor fetal growth). The study (Ref # 27) by Bi WG is real science. Also, what is interesting is that in the observational studies more significant findings were found which (I think) means that observational studies are done on different populations by different types of investigators. It was interesting that in the observation studies lower vitamin D levels were associated with SGA but NOT <2500 grams. Low Birth Weight (<2500grams is not a good index and percentiles should be used because percentiles by GA can give a spread (<10th) but also >10 to zero in on what is important between 10 and 25 as an example.
REVIEWER #3 REPLY: We are pleased that you find our conclusions plausible.

REVIEWER #3 COMMENT: The authors mention the “Acute Phase Reaction” several times. There are many examples of the acute phase reaction. One of the most important APRs is related to vitamin D. During an infection Toll receptor on a macrophage are stimulated by bacterial products which sends a signal into the nucleus to turn on a gene to make hydroxylase to add a OH ion to circulating vitamin D (providing there is no vitamin D deficiency) to produce active form of vitamin D [1-25 (OH)2 D] which goes back into the nucleus to turn on a second gene to make Cathelicidin that kills Tuberculosis Bacterium [that is why people went to Sanitoriums] to recover from TB. Today the issue of vitamin D studies is that there are two factors important as a cause of diseases. #1) It’s a combination of factors and the future of vitamin D studies to assess its importance will require a NEW

APPROACH TO STUDIES which will require a systems biology approach. #2) Vitamin D and its role in health is undergoing dramatic change because of the issues of metabolism of vitamin D. Is it synthesis, metabolism, the role of binding protein or is it receptor genetics?
REVIEWER #3 REPLY: Yes, it will be very interesting to continue to follow the Vitamin D science.

REVIEWER #1 COMMENT: No suggestions other than to check the median and range values reported for AMSTAR ratings (median 8, range 8-11). Mean instead of median?
REVIEWER #1 REPLY: We checked these numbers and the median is 8 as most of the studies had a score of 8 with only a few scoring above this. The mean would be 8.76.