PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (**http://bmjopen.bmj.com/site/about/resources/checklist.pdf**) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

ARTICLE DETAILS

TITLE (PROVISIONAL)	Modifying Maternal Sleep Position in the Third Trimester of Pregnancy with Positional Therapy: A Randomized Pilot Trial
AUTHORS	Kember, Allan; Scott, Heather; O'Brien, LM; Borazjani, Ali; Butler, Michael; Wells, Jesse; Isaac, Andre; Chu, Kaishin; Coleman, Jerry; Morrison, Debra

VERSION 1 – REVIEW

REVIEWER	Alessio Signori University of Genoa, Italy
REVIEW RETURNED	13-Nov-2017

GENERAL COMMENTS	None

REVIEWER	Timor Faber Erasmus University Medical Center, The
REVIEW RETURNED	05-Dec-2017

GENERAL COMMENTS	Thank you for the opportunity to review this interesting manuscript. The authors have conducted a double-blind randomized sham-controlled crossover trial to determine the differences in the mean proportions of supine sleep between pregnant women using a positional therapy device and those using a sham. Whereas the authors have clearly tried to use the best available (gold-standard) methods to conduct their trial, there are a few issues that are concerning about this manuscript. First, the results from the data analysis should be interpreted more carefully. Even though the authors rightfully mention that “caution should be taken when extrapolating the results to the home setting and across the third trimester” on page 3, I believe they are not cautious enough in their statements regarding the results of the trial throughout the abstract and the rest of the manuscript. For example, the statement “... and has important implications for the prevention of late unexplained stillbirth” in the abstract is based on unjustified extrapolation of the results. In general, I would like to advise the authors to be more subtle in their wording throughout the manuscript, so that their statements reflect proper interpretation of the results. For example, the authors should mention that they were unable to demonstrate any differences in sleep architecture and duration instead of stating that their analysis demonstrated no difference (page 18). Second, the reporting of the linear mixed effects models lacks detail which makes it difficult for the reader to decide whether proper

	analyses have been conducted. For example, it is unclear to me why the authors decided to report differences in means when the mean proportions between the two groups had unusually large standard deviations. Also, why did the authors report the means and standard deviations of the primary outcome and some secondary outcomes (time supine; sleep latency; proportion stage 1 sleep; proportion stage 3 sleep; and AHI) when their standard deviations are also unusually large and therefore probably not normally distributed. It would be helpful if the authors would report the median and interquartile range for these secondary outcomes, and elaborate on their choice of linear mixed models for these analyses. Third, I would like the authors to discuss the statistical limitations of their analyses in more detail. For example, the authors mention that the study may be underpowered, but do not elaborate enough on how this could affect their findings. This is an important issue that demands more attention. Last, I would like to suggest some minor changes that could improve the manuscript. - Please insert the minus sign before 0.3 on page 13 first paragraph. - The term percentage point change is commonly used for absolute differences in percentages, whereas percentage change refers to relative differences in percentages. - If possible, please provide the raw data and data analysis script as an attachment. This paper needs some minor revision that would better reflect a conclusion that is based on proper interpretation of the data. Nonetheless, I commend the authors for having conducted an important trial that addresses supine sleeping in pregnant women.

REVIEWER	Safwaan Adam University of Manchester, United Kingdom, Manchester University NHS Foundation Trust, United Kingdom, Salford Royal NHS Foundation Trust, United Kingdom
REVIEW RETURNED	09-Dec-2017

GENERAL COMMENTS	I was asked to focus solely on the statistical methods used and I felt that these were appropriate. The methods were well-described and the results pertaining to these presented clearly in my opinion.

REVIEWER	Marie Gantz RTI International, USA
REVIEW RETURNED	18-Jan-2018

GENERAL COMMENTS	Overall: This is a very interesting study with an appropriate study design. However, my major concerns are with discrepancies between the protocol and manuscript as described below. In addition, more information should be presented regarding the crossover design and the potential for period and crossover effects (comment 16). 1) Page 6, line 30: Please explain why a randomized trial involving pregnant women was not subject to oversight by an independent data and safety monitoring committee. 2) Page 9, line 6: Please explain whether the power and sample size

	analysis accounted for the crossover design. 3) Primary outcome: According to the protocol, the first primary outcome of the study was percent of time spent in the supine and right-lateral positions. In the manuscript, only the supine position is reported on. The manuscript should be consistent with the protocol. 4) Primary outcome: According to the protocol, the second primary outcome of the study was "PrenaBelt User Feedback Questionnaires." The methods, results, and discussion do not present this as a primary outcome. 5) Page 9, line 8: Please explain mean proportion of time spent sleeping supine (and right lateral) was selected as the primary outcome as opposed to total length of time. Intuitively, length of time seems more relevant with respect to risk of adverse effects of sleeping supine. 6) Page 9, line 37: The manuscript states "An envelope was drawn at random..." A typical randomization scheme would be to randomly order the envelopes and then open them consecutively (in that random order). Please clarify how randomization of the envelopes was performed. 7) Page 10, line 13: For a simple balanced crossover design such as this, the typical analysis approach would be to test whether the difference in the effect of treatment A and treatment B on the same participant is different from zero using a paired t test or Wilcoxon rank sum test. Paired t tests and Wilcoxon signed rank tests are the analysis methods described in the study protocol. However, the analysis of the primary outcome in the manuscript differs from this approach. The planned analysis methods from the protocol should be used for the primary outcome as they are for the secondary outcomes. 8) Page 10, line 13: The power analysis indicates use of a one-sided test, but the primary outcome analysis appears to have used a two-sided test. Please reconcile this discrepancy. 9) Page 11, line 15: The results indicate that 25 participants was the planned sample size, but the sample size description indicates that 25 pairs were required for adequate power. Please reconcile this discrepancy. 10) Page 11, Sample Characteristics: Please provide information about the length of time between sleep studies for the same participant. 11) Page 12, Table 1: Assuming that the 3 participants who did not complete both sleep studies were excluded from analysis, the more relevant population for Table 1 is the 20 who were analyzed. The authors should also include differences between those who did and did not complete the study. 12) Page 16, line 26: Please report whether the primary outcome was normally distributed, since this has potential impact the appropriateness of the analysis method. 13) Results: Please also report the length of time spent supine while sleeping by treatment. 14) Page 20, line 55: What is the reference for the statement that "most pregnant women continue to spend a significant amount of time supine during sleep in late pregnancy?" 15) Page 21, line 8: Please clarify whether the statement that "pregnant women's estimates of time in each sleep position...underestimate the time they spend supine" is based on the current study (which may not be generalizable due to the small sample size) or provide a reference. 16) The potential for period effects in this crossover study is alluded to in the discussion of study strengths and weaknesses. However, it warrants more attention in the methods, results, and discussion give

	that this is a crossover design study. In this study, one could expect period effects due to a variety of factors including increased patient knowledge and comfort with the process, and increased discomfort sleeping due to being further along in pregnancy. Similarly, potential carryover effects and the length of the washout period should be explicitly addressed in the methods and results sections (and discussion if appropriate). 17) Consort: Diagram should provide reasons for the 3 dropouts.

REVIEWER	Margaret Miller Warren Alpert Medical School, Brown University, Women's Medicine Collaborative, Providence, RI
REVIEW RETURNED	24-Jan-2018

GENERAL COMMENTS	The study was well designed and outcomes clearly identified. As the authors point out, the sample size is smaller than planned and may not have been adequately powered to detect significant differences. While this may be a significant limitation, this study provides a strong basis for future research in this area. Also agree with authors that the study would be more robust if baseline sleep data was also included.

VERSION 1 – AUTHOR RESPONSE

ID	Comment/Request	Response
1	Please do not start your abstract with the trial design section. Please reformat your abstract according the sub-headings given in BMJ Open's instructions for authors	We have reformatted our abstract according to the sub-headings given in BMJ Open's instructions for authors, "A structured abstract" section.
2	The methods section describes the primary outcome, but there appears to be a lot of other outcomes in the trial registry that are not reported. Please report on all outcomes in the methods and results sections.	The trial registry for our study (NCT02377817) includes: • One "Primary Outcome Measure": ○ Body Position (*Proportion of sleeping time spent in the supine, prone, left-lateral, and right-lateral positions for each participant*) ▪ In our original manuscript, we specified and reported the primary outcome as "percentage of sleep time supine" in **Table 2**. ▪ The percentage of sleeping time spent in the prone, left-lateral, and right-lateral positions were not reported in our original manuscript; therefore, we added these data to **Table 3** as secondary outcomes. We also added the following sentence to **RESULTS** > **Secondary Outcomes**: • "*The mean (SD) percentage of time spent sleeping in the left-lateral and right-lateral positions was 54.6% (26.7) and 30.3% (27.8), respectively. No time was spent sleeping prone.*" • One "Secondary Outcome Measure": ○ User Feedback (*Each participant will complete the PrenaBelt User Feedback Questionnaire, which is designed to gather feedback on how she slept, her experience using the PrenaBelt, and her input on how the PrenaBelt could be*

ID	Comment/Request	Response
		improved.) ▪ In our original manuscript, we reported the participants' experience regarding how she slept (sleep onset position, waking position, number of position changes, percent/proportion of time in each position) in **Table 4**. We also reported the participants' experience using the PrenaBelt (satisfaction, comfort level, and intention to continue using the PrenaBelt) in **Table 3**. ▪ The participant's input on how the PrenaBelt could be improved was not reported in our original manuscript; therefore, we added the following sentence to **RESULTS** > **Secondary Outcomes**: • "*Six participants (30%) indicated that they would make changes to the PrenaBelt – of these, five were with regard to comfort and one was with regard to sizing*." • Ten "Other Outcome Measures": o Apnea-Hypopnea Index (AHI) (*Will be reported as a total AHI as well as AHI while supine and AHI while non-supine.*) ▪ In our original manuscript, we reported the total AHI in **Table 3**. ▪ The sleep reports generated by our sleep diagnostic software (Embla Sandman Elite) were not configured to compute a supine AHI and non-supine AHI, so we are unable to report these values; however, the software did compute a supine RDI and non-supine RDI (see "RDI" below). The RDI is the sum of the AHI (apneas and hypopneas) and RERA index (respiratory-event related arousals). o Blood Oxygen Saturation (SaO2) (*Mean SaO2, Min SaO2, and Max SaO2 during Awake, Rapid Eye Movement (REM), and Non-REM (NREM) states*) ▪ Note that we use the term "SpO2", not "SaO2". "SpO2" indicates that the blood oxygen saturation was measured by pulse oximetry (non-invasive), whereas "SaO2" indicates direct measurement via arterial blood gasses (invasive). In our study, we used pulse oximetry. Reference to "SaO2" is an error in the trial registry, which has been corrected. ▪ In our original manuscript, we reported the Mean, Minimum, and Maximum SpO2 during sleep. We edited the **RESULTS** > **Secondary Outcomes** and **Table 3** to remove this measure and replace it with the Mean, Minimum, and Maximum SpO2 during REM and NREM sleep states. ▪ Note that we did not report SpO2 values during "Awake" state because these values are prone to artifact when donning and doffing the SpO2 fingertip probe at the beginning of the sleep study and before/after bathroom breaks, which can be difficult to ascertain and remove from the dataset. Also, since this study is a sleep study, we are more interested in the SpO2 during sleep states (REM and NREM), which we have reported. o Heart Rate (HR) (*maternal Mean HR, Min HR, Max HR during Awake, REM, and NREM states*) ▪ The HR was not reported in our original or revised manuscript. ▪ Justification for not reporting HR: Although we recorded electrocardiography (ECG), the sleep reports generated by our sleep diagnostic software were not configured to compute HR from the ECG, so we are unable to report these values. o Total Sleep Time (TST)

ID	Comment/Request	Response
		▪ In our original manuscript, we reported TST in **Table 3**. ○ Snoring (presence of) ▪ Snoring was not reported in our original manuscript; therefore, we added snoring data to **Table 3**. We also added the following sentence to **RESULTS** > **Secondary Outcomes**: • "*Presence of snoring was detected on 26 of 40 (65%) nights*." ○ Respiratory Disturbance Index (RDI) (*respiratory effort-related arousal index (RERA index), total RDI, as well as RDI while supine and RDI while non-supine*) ▪ In our original manuscript, we reported RERA index in **Table 3**. ▪ The total RDI, supine RDI, and non-supine RDI was not reported in our original manuscript; therefore, we added these data to **Table 3**. We also added the following sentence to **RESULTS** > **Secondary Outcomes**: • "*…median (IQR)… respiratory disturbance index (RDI) was 0.6 (0-1.6), RDI supine was 0 (0-0.9), RDI non-supine was 0.5 (0-1.6)…*" ○ Sleep latency ▪ In our original manuscript, we reported Sleep latency in **Table 3**. ○ Sleep efficiency ▪ In our original manuscript, we reported Sleep efficiency in **Table 3**. ○ Total arousals (*total arousal index, is further classified as spontaneous, periodic leg movement, or respiratory arousal index)* ▪ In our original manuscript, we reported the number of sleep stage shifts and the number of awakenings in **Table 3**; however, arousals are the metrics that are more typically reported in the sleep literature compared to stage shifts and awakenings. The total arousal index, spontaneous arousal index, periodic leg movement arousal index, and respiratory arousal index were not reported in our original manuscript; therefore, we edited the manuscript to *remove* the sleep stage shifts and awakenings from **Table 3** and **RESULTS** and *add* arousals to **Table 3**. We also added the following sentence to **RESULTS** > **Secondary Outcomes**: • "*The median (IQR) total arousal index was 11.3 (8.4-18.0), spontaneous arousal index 9.8 (7.0-12.0), periodic leg movement arousal index 0 (0-0.8), and respiratory arousal index 0.4 (0-1.0).*" ○ Number of position changes ▪ In our original manuscript, we reported the number of position changes in **Table 4**.
3	How does the belt change how a person sleeps? Is there evidence that "the balls apply pressure points across the user's lower back, prompting her to reposition herself in a lateral position to maintain comfort"?	The PrenaBelt device is predicated on the "tennis-ball technique" (TBT) method of positional therapy (PT), as such, this is the *theoretical* mechanism of action of the PrenaBelt. The TBT method delivers sensory feedback to the supine user in the form of pressure and pain sensation via activation of mechanoreceptors in the skin and soft tissue overlying the pressure point, and eventually the user readjusts his/her position to relieve this noxious stimulus and regain comfort. Since the 1980's [*Patient's wife cures his snoring.* Chest. 1984;85:582.], PT has been formally studied and employed in the

ID	Comment/Request	Response
		field of sleep medicine as a simple, safe, and effective treatment to help individuals with sleep-disordered breathing avoid the supine sleeping position and maintain a lateral-sleeping position (see **INTRODUCTION – Background**). The TBT method is just one of several methods of PT, e.g., pressure points, auditory, and/or vibratory alarms (see **DISCUSSION – Strengths and Weaknesses in Relation to Other Studies**). • This study was completed to study the PrenaBelt. Since there is no evidence that "the balls apply pressure points across the user's lower back, prompting her to reposition herself in a lateral position to maintain comfort", we have edited the manuscript to indicate that this is the *theoretical* mechanism of action of the PrenaBelt based on the TBT method of PT and contend that this postulation is justified based on: o Predication on the TBT method of PT (PMC4298771, PMC2883034, PMID 9400908) o Physiology of mechanoreceptors (see also PMID 13891095).
4	How did you select your sample size? What was this based on? How did you know that the trial had sufficient power to measure what you set out to find?	Selection of our sample size was addressed in our original manuscript submitted to BMJ Open (see **METHODS – Sample Size**) and our Research Protocol (see **1.4.1.5 – Sample Size**) where we acknowledged that there was no preliminary data on which to base our sample size calculation. As a rule of thumb, we powered our study to enable a detectable effect (d) of -0.5, which is a medium effect size per the literature regarding Cohen. This effect size would allow us to detect a difference of half a standard deviation in the mean proportion of time spent sleeping supine on the sham night versus the treatment night. Per common convention, we set our power (β) to be 0.80 and significance level (α) to be 0.05. With these parameters, the sample size required for a one-sided paired t-test is n=25. As noted in our original manuscript (see **RESULTS**), "*The originally planned sample size of 25 participants was not reached due to unforeseen budget restrictions preventing recruitment beyond 20 participants*." When the decision was made to reduce the sample size from 25 to 20 (on 25OCT2016), we conducted an interim analysis (unplanned) with n=13 to ensure our trial would still have sufficient power at n=20. While we were aware that interim data may not be reflective of the eventual, fully-powered findings, our interim analysis predicted sufficient power (0.91) with n=20.
5	What is the clinical significance of how women sleep on one night in their pregnancy?	This was a pilot trial to determine whether maternal sleep position was modifiable. As such, we only tested the PrenaBelt for one night of sleep to investigate the effect of the device on maternal sleep parameters, particularly sleep position. Knowing whether a device such as the PrenaBelt can achieve a change in sleep behavior is critical before we can launch a large longitudinal study. Emerging data demonstrates the negative impact of even one night of supine sleep on the fetus.[1,2] From clinical practice, we know that maternal posture influences hemodynamics during labor; changing

ID	Comment/Request	Response
		position from supine to left lateral recumbent increases cardiac output and can often mitigate fetal distress.[3] Thus, with a growing literature suggesting that supine sleep position is a risk for stillbirth,[4-8] our findings that maternal sleep position can be modified (and supine sleep time reduced) has potentially significant impact that warrants further investigation with longitudinal studies. • [1] Stone PR, et. al., *Effect of maternal position on fetal behavioural state and heart rate variability in healthy late gestation pregnancy*. J Physiol. 2017 Feb 15;595(4):1213-1221. • [2] Warland J, Dorrian J, Kember A, Phillips C, Morrison J, O'Brien L, Borazjani A. *Modifying Maternal Sleeping Positions In Late Pregnancy Through Positional Therapy: A Feasibility Study*. Under review by the Journal of Clinical Sleep Medicine. • [3] Thurlow JA, Kinsella SM. *Intrauterine resuscitation: active management of fetal distress*. Int J Obstet Anesth. 2002 Apr;11(2):105-16. • [4] Stacey T, et al. *Association between maternal sleep practices and risk of late stillbirth: a case-control study*. BMJ. 2011;342:d3403. • [5] Owusu JT, et al. *Association of maternal sleep practices with pre-eclampsia, low birth weight, and stillbirth among Ghanaian women.* Int J Gynaecol Obstet. 2013 Jun;121(3):261–5 • [6] Gordon A, et al. *Sleep position, fetal growth restriction, and late-pregnancy stillbirth: the Sydney stillbirth study*. Obstet Gynecol. 2015 Feb;125(2):347–55. • [7] McCowan LME, et al. *Going to sleep in the supine position is a modifiable risk factor for late pregnancy stillbirth; Findings from the New Zealand multicentre stillbirth case-control study*. PLoS One [Internet]. 2017;12(6):e0179396. • [8] Heazell AEP, et al. *Association between maternal sleep practices and late stillbirth – findings from a stillbirth case-control study*. BJOG 2018; 125:254–262.
6	We are concerned that these findings are a little preliminary and that this might be better described as a pilot on which a further study could be planned. Throughout the paper please tone down your language and do not overstate your findings.	Based on a review of definitions given by the National Institute for Health Research Trials and Studies Coordinating Centre, our study is best described as a pilot study. Further clarification is given by Arnold et al [1] who define a pilot trial as a stand-alone pilot study that includes a randomization procedure. • [1] Arnold DM, Burns KE, Adhikari NK, Kho ME, Meade MO, Cook DJ: McMaster Critical Care Interest Group. *The design and interpretation of pilot trials in clinical research in critical care*. Crit Care Med 2009, 37(Suppl 1):S69-74. We have edited the title and manuscript throughout to indicate that our study is a report of a *pilot trial* and adjusted the statement of our findings accordingly. Changes: • <u>Title</u>: Modifying Maternal Sleep Position in the Third Trimester of Pregnancy with Positional Therapy: A Pilot Trial • Deleted from Abstract > **CONCLUSIONS**: "*and has important implications for the prevention of late unexplained stillbirth and for the design of a public health campaign about safe sleep position in late pregnancy*."

ID	Comment/Request	Response
		• Prefaced "**Trial Design**" subheading (see **METHODS**) with "**Pilot**". • Indicated "this was a pilot trial" in **DISCUSSION** > **Strengths and Weaknesses Relative to Other Studies**. • We removed "*which could have significant clinical implications if supine sleep needs to be avoided and self-reports are relied upon by maternity care providers for risk stratification*" from **DISCUSSION** > **Strengths and Weaknesses Relative to Other Studies** as this language may be viewed out of proportion to the strength of our findings and is addressed in a lighter tone in **DISCUSSION** > **Meaning of the Study**. • We removed references to a "*public health education campaign*" throughout our manuscript as such campaigns would ideally be based upon stronger evidence than that from pilot trials like ours. • We changed "robust design" to read "rigorous methodology" in **DISCUSSION** > **Strengths**, which is appropriate terminology to describe a pilot trial per reference [1] above. • We changed the last sentence of the **DISCUSSION** > **Future Research** from "*Large, multi-ethnic studies that include women with a range of pregnancy and health conditions are imperative to refute or confirm the findings*." to read, "*The results of our pilot trial warrant future, large, multi-ethnic studies that include women with a range of pregnancy and health conditions to ascertain if the observed effects persist.*"
7	Please provide a more thorough discussion of the study's limitations in the discussion section.	Noted. See response to Reviewer #2, Row ID #9 and #10, and Reviewer #4, Row ID #13 and #27 (below).
8	First, the results from the data analysis should be interpreted more carefully. Even though the authors rightfully mention that "caution should be taken when extrapolating the results to the home setting and across the third trimester" on page 3, I believe they are not cautious enough in their statements regarding the results of the trial throughout the abstract and the rest of the manuscript. For example, the statement "... and has important implications for the prevention of late unexplained stillbirth" in the abstract is based on unjustified extrapolation of the results. In general, I would like to advise the authors to be more subtle in	We have edited the manuscript throughout to be subtler and cautious in the statement of our findings in order to reflect a balanced and proper interpretation of the results. See response to Editor comment, Row ID #6 (above). In addition, we specifically addressed the example given by the reviewer regarding sleep architecture in: **DISCUSSION** > **Principal Findings** – we reworded the sentence to read: • "*Use of the PrenaBelt resulted in a 6.8% absolute reduction (38% relative reduction, 24.4 minutes) in the mean percentage of sleep time supine in comparison with the sham, and we were unable to demonstrate an effect on maternal sleep architecture or respiration*" **DISCUSSION** > **Strengths and Weaknesses in Relation to Other Studies** – we reworded the sentence to read: • "*In our analysis, we were unable to demonstrate any differences in sleep architecture or duration despite a significant difference in body position (less supine time)*."

ID	Comment/Request	Response
	their wording throughout the manuscript, so that their statements reflect proper interpretation of the results. For example, the authors should mention that they were unable to demonstrate any differences in sleep architecture and duration instead of stating that their analysis demonstrated no difference (page 18).	
9	Second, the reporting of the linear mixed effects models lacks detail which makes it difficult for the reader to decide whether proper analyses have been conducted. For example, it is unclear to me why the authors decided to report differences in means when the mean proportions between the two groups had unusually large standard deviations.	The linear mixed effects (LME) model does not depend on the underlying normality of the outcome (note that our primary outcome was not normally distributed – see response to Reviewer #4, Row ID #23). The LME model depends on the normality of the errors. We acknowledge that the two groups had large standard deviations, which reflects the large variability that would be expected when measuring percentage of time spent supine in a relatively small sample. When the one-sided paired Wilcoxon rank sum test is performed on the primary outcome per the original planned analysis in the Research Protocol (as suggested by Reviewer #4), the results of our study do not change (p-value on LME model analysis = 0.04, and p-value on one-sided paired Wilcoxon rank sum test = 0.03). Given that the LME p-value is 0.04 (0.03 on Wilcoxon) for the difference in our primary outcome ("Percent of sleep time supine (%)"), this outcome is especially fragile,[1] i.e., a small change in the outcome for a few participants shift the overall results from being statistically significant to non-significant. To this end, we have added the following sentence (see **DISCUSSION** > **Weaknesses**) to indicate that this finding is fragile: • *Because of the fragility of our primary outcome (p 0.04), a confounding could shift our conclusions into statistical non-significance. This stresses the importance of future research to ascertain if our observed effects persist.* • [1] Phil Davis, Michael Butler, Kirk Magee, Chris Nickson, Seth Trueger. *How Robust Are Studies in the American Board of Emergency Medicine Maintenance of Certification Lifelong Learning and Self-assessment? An Examination of Fragility and Bias of Included Randomized Controlled Trials*. **Academic Emergency Medicine: Education and Training**; Volume 1, Issue 4, October 2017, Pages 280–286. We have reviewed the normality testing of the primary and secondary outcomes. Our primary outcome was not normally distributed. Some of our secondary outcomes were normally distributed and some were not. We have edited the manuscript to

ID	Comment/Request	Response
	Also, why did the authors report the means and standard deviations of the primary outcome and some secondary outcomes (time supine; sleep latency; proportion stage 1 sleep; proportion stage 3 sleep; and AHI) when their standard deviations are also unusually large and therefore probably not normally distributed. It would be helpful if the authors would report the median and interquartile range for these secondary outcomes, and elaborate on their choice of linear mixed models for these analyses.	report the median and IQR for non-normally distributed continuous variable outcomes instead of the mean and standard deviation. We added a statement below **Table 1** and **Table 3**: • "*Normally distributed continuous variables are reported as mean (SD), and paired t-test is used to test for difference and indicated by ‡. Non-normally distributed continuous variables and discrete data (Satisfaction, Comfort, and Intention to use) are presented as median (IQR), and two-sided paired Wilcoxon signed rank test is used to test for difference and indicated by †. Count data (Presence of snoring) are presented as frequency (%).*" Please note that the LME model was <u>not</u> used for any of the secondary outcomes except the "Time supine (minutes)" (see "*" in **Table 3**), which is just the absolute value in minutes of primary outcome.
10	Third, I would like the authors to discuss the statistical limitations of their analyses in more detail. For example, the authors mention that the study may be underpowered, but do not elaborate enough on how this could affect their findings. This is an important issue that demands more attention.	Noted. See response to Reviewer #2, Row ID #9 regarding fragility (above). We found that the PrenaBelt reduced the median percentage of supine sleep by 13%, i.e., from 16.4% (56.8 minutes) to 3.5% (12.3 minutes) – avoidance of about 45 minutes supine. Due to our small sample size, we acknowledge that our findings may overstate the true reduction in percent of supine sleep. However, it is important to note that the median sleep time in our study was relatively short (352 minutes = 5.9 hours), which is not uncommon in laboratory sleep studies due to factors such as unfamiliar sleep environment and logistics. In the home, the sleep time is likely to be significantly longer, e.g., our participants reported a median of 8 hours overnight sleep duration. If our results hold in the home environment, a 13% reduction in the median percentage of supine sleep translates to avoidance of 60 minutes of supine sleep based on an 8 hour sleep time. Even if our results are overstated and the reduction is less than we observed, the PrenaBelt is still likely to have a considerable impact on the minutes of supine sleep avoided (e.g., a 7% reduction in the median percentage of supine sleep translates to avoidance of 33 minutes of supine sleep based on an 8 hour sleep time). We added the following sentence to the **DISCUSSION** > **Weaknesses**: • "*Due to the small sample size, this study may be underpowered and potentially overstate the true reduction in percentage of*

ID	Comment/Request	Response
		supine sleep; however, the sleep time in our study was relatively short (median 5.87 hours). Our participants reported a median of 8 hours overnight sleep duration at home; therefore, even if the reduction in supine sleep is lower than we observed, over a longer sleep duration, the PrenaBelt is still likely to result in a significant amount of supine sleep avoided."
11	Last, I would like to suggest some minor changes that could improve the manuscript. (1) Please insert the minus sign before 0.3 on page 13 first paragraph. (2) The term percentage point change is commonly used for absolute differences in percentages, whereas percentage change refers to relative differences in percentages. (3) If possible, please provide the raw data and data analysis script as an attachment.	Noted. (1) Done. (2) Thank you for describing this terminology; however, we feel that the use of the terms "absolute" and "relative" will convey the clearest meaning for the readers as not many will be familiar with the subtle but important difference between the terms "percentage point change" and "percentage change". (3) We have attached our data analysis scripts and output as supplementary files to our manuscript and noted this in the body of the manuscript (see **METHODS** > **Statistical Methods** and **OTHER INFORMATION** > **Data Sharing**). With regard to our raw data, we are currently seeking guidance from the IWK Research Ethics Board (Halifax, Canada) to share our raw data via Dryad per BMJ Open's preference; however, sharing our raw data may not be possible because raw data sharing was not directly addressed in our consent form, and the sample size was only 20 participants, which yields a higher risk of re-identification of the de-identified data. Therefore, we removed the following sentence from **OTHER INFORMATION** > **Data Sharing**: "*Complete dataset is available from the Dryad repository, DOI:*"
12	1) Page 6, line 30: Please explain why a randomized trial involving pregnant women was not subject to oversight by an independent data and safety monitoring committee.	The Research Ethics Board responsible for this study, IWK REB, gave ethics approval of the proposed trial on 16JUN2015 (Project #1018753). This included approval of our safety monitoring plan and follow-up care plan, which was detailed in our Ethics Approval Submission (EAS) Form submitted to the IWK REB and was outlined in **Section 1.6 – Definition of Adverse Events** and **Section 2.4 – Minimization of Potential Harms** in our Research Protocol. The EAS Form specifically asked, "Does the study have an independent data and safety monitoring board?", to which we answered "No" and cited the reasons it was not necessary per Sections 1.6 and 2.4 in our Research Protocol, namely: • The PrenaBelt device is a non-invasive medical device of Class I designation (see attached letter from Health Canada – for Editors only). • Pregnant women typically sleep with many pillows supporting their body, including a pillow behind their back to avoid the supine position. The PrenaBelt is a positional therapy device that may assist pregnant women to avoid supine sleep. Positional therapy devices have been shown to be safe and approved for use by humans by the US Food and Drug Administration [1]. In addition, maternal body pillows, regular pillows, and pelvic belts (lumbar support) have been used by pregnant women during sleep without reports of serious adverse effects for the mother or

ID	Comment/Request	Response
		neonate [2]. ○ [1] U.S. Food and Drug Administration. (2010, May) US Department of Health and Human Services. K100160 ○ [2] Victoria Pennick and Sarah D Liddle, "Interventions for preventing and treating pelvic and back pain in pregnancy (Review)," The Cochrane Collaboration, London, Review 2013. • Polysomnography is a non-invasive, routine sleep diagnostic test. As such, participants in this study were at no greater risk of harms when completing the activities of this study than those risks they encounter in their everyday life. • Eligibility criteria for inclusion into this study was such that participants were inherently within the lowest risk stratification of pregnant women.
13	2) Page 9, line 6: Please explain whether the power and sample size analysis accounted for the crossover design.	The power and sample size analysis did not account for the crossover design. Our power and sample size analysis was via a one-sided t-test. To account for the crossover design in our power and sample size analysis, simulation techniques would be required and we would need prior knowledge about the variables in order to determine effect; however, given the preliminary nature of this study, we had no prior knowledge of variables required for such a simulation. To this end, we have added a sentence to indicate this limitation (see **DISCUSSION** > **Weaknesses**): • "*The power analysis was performed using the one-sided t-test, which could be a source of systemic confounding.*"
14	3) Primary outcome: According to the protocol, the first primary outcome of the study was percent of time spent in the supine and right-lateral positions. In the manuscript, only the supine position is reported on. The manuscript should be consistent with the protocol.	This is correct per our Research Protocol Revision V.2015/01/26; however, note that in the Research Protocol **Section 1.1.3 Background – Maternal Device**, we made the following provision: • "*The PrenaBelt is also designed for adjustability and comfort. As such, its configuration can be easily adapted by the user to only help her avoid sleeping on her back if she requires the option of sleeping on her right side as well as her left for comfort reasons."* Many physiologic studies and a recent systematic review [1] confirm that when a pregnant woman assumes the supine or right-sided position, compared with the left-sided position there is decreased blood return to the maternal heart, decrease maternal cardiac output, and decreased fetal oxygenation. Evidence for increased association between supine sleeping position in late pregnancy and late stillbirth is most consistent and robust.[2,3-5] Right-lateral sleeping position has been implicated too in the earliest study of this association,[2] but this finding has not been reproduced in later, larger studies.[3-5] At the time we wrote our research protocol, only reference [2] was published, so we included right-lateral sleep as part of the primary outcome but made provision for excluding it based on our Research Ethics Board's (IWK REB) recommendation and participant comfort.[6] Later and larger studies [3-5] did not corroborate this inclusion and all participants in our study opted to have freedom to sleep on their right; therefore, to avoid confusion for our readers, we contend that the primary outcome should solely

ID	Comment/Request	Response
		be “Percent of sleep time supine (%)”, which is consistent with the Research Protocol given the provision that was made and deemed acceptable by the IWK REB at the time of approval. • [1] Cluver C, et al. *Maternal position during caesarean section for preventing maternal and neonatal complications.* The Cochrane Database of Systematic Reviews 2010, Issue 6. Art. No.: CD007623 • [2] Stacey T, et al. *Association between maternal sleep practices and risk of late stillbirth: a case-control study*. BMJ. vol. 342, 2011. • [3] Gordon A, et al. *Sleep Position, Fetal Growth Restriction, and Late-Pregnancy Stillbirth: The Sydney Stillbirth Study.* Obstet Gynecol, vol. ePub, January 2015. • [4] McCowan LME, et al. *Going to sleep in the supine position is a modifiable risk factor for late pregnancy stillbirth; Findings from the New Zealand multicentre stillbirth case-control study*. PLoS One. 2017;12(6):e0179396. • [5] Heazell AEP, et al. *Association between maternal sleep practices and late stillbirth – findings from a stillbirth case-control study*. BJOG 2018; 125:254–262. • [6] Warland J, Dorrian J *Accuracy of Self-Reported Sleep Position in Late Pregnancy*. PLoS One. 2014 Dec 23;9(12):e115760
15	4) Primary outcome: According to the protocol, the second primary outcome of the study was "PrenaBelt User Feedback Questionnaires." The methods, results, and discussion do not present this as a primary outcome.	This is an oversight on our part in the language of our Research Protocol. The “PrenaBelt User Feedback Questionnaire” data are secondary outcomes (see the Outcome Measures section in our trial registration). See response to Editor comment, Row ID #2 (above) regarding User Feedback.
16	5) Page 9, line 8: Please explain mean proportion of time spent sleeping supine (and right lateral) was selected as the primary outcome as opposed to total length of time. Intuitively, length of time seems more relevant with respect to risk of adverse effects of sleeping supine.	Please note that we presented and analyzed *both* the mean proportion of time spent sleeping supine (**Table 2**, and **RESULTS** > **Primary Outcome**) and mean length of time spent sleeping supine (**Table 3**, and **RESULTS** > **Secondary Outcomes**) in the manuscript. We used the same statistical methods used for the primary outcome (see response to Reviewer #2, Row ID #9). While we agree that the length of time sleeping supine seems more relevant with respect to risk of adverse effects of sleeping supine, we currently do not know how much time (quantitatively) spent sleeping supine is clinically significant (see response to Editor comment, Row ID #5). The reason that the percentage of time sleeping supine was chosen as the primary outcome is because our study was designed to test the PrenaBelt – not the adverse effects of sleeping supine. The reduction in the absolute length of time spent sleeping supine is not

ID	Comment/Request	Response
		useful in evaluating the PrenaBelt without the context of the total time spent asleep, which the percentage inherently incorporates. For example, a 30-minute reduction in time supine is a 100% reduction if the total supine time was 30 minutes, but it is only a 33% reduction if the total supine time was 90 minutes. Therefore, the impact of the PrenaBelt on minimizing supine sleep is best understood when the percentage reduction is considered, and this understanding is enhanced with addition of the absolute reduction in minutes, which we presented as a secondary outcome.
17	6) Page 9, line 37: The manuscript states "An envelope was drawn at random..." A typical randomization scheme would be to randomly order the envelopes and then open them consecutively (in that random order). Please clarify how randomization of the envelopes was performed.	This is an error in the language of our original manuscript due to a misunderstanding by an author (AK). The randomization was by virtue of the random allocation sequence generated by the computer and concealed by the independent statistician (MB) into unmarked, security-tinted, sealed envelopes. These envelopes were then given (by MB) to the recruiter and placed in a desk drawer in a stack. When a participant was ready to be randomized, the envelope at the top of the stack was drawn and opened; therefore, the envelopes were not actually drawn at random because the randomization had already been completed and concealed by the statistician. We have changed the language of the manuscript to replace “at random” with “in sequence”.
18	7) Page 10, line 13: For a simple balanced crossover design such as this, the typical analysis approach would be to test whether the difference in the effect of treatment A and treatment B on the same participant is different from zero using a paired t test or Wilcoxon rank sum test. Paired t tests and Wilcoxon signed rank tests are the analysis methods described in the study protocol. However, the analysis of the primary outcome in the manuscript differs from this approach. The planned analysis methods from the protocol should be used for the primary outcome as they are for the secondary outcomes.	Noted. Please see response to Reviewer #2, Row ID #9 regarding the appropriateness of a linear mixed effects (LME) model analysis in the context of within-subject dependency. We respectfully contend that the LME analysis is the appropriate analysis in context of random effects of within-subject differences and study night (night 1 vs. night 2). Further, Reviewer #3 was asked to focus solely on the statistical methods used and results pertaining to these and felt that they were appropriate. Finally, our results do not change when the planned analysis per the Research Protocol is completed. The primary outcome (“Percent of sleep time supine (%)”) did not follow a normal distribution. When the one-sided paired Wilcoxon rank sum test is performed on the primary outcome per the original planned analysis in the Research Protocol (as suggested by Reviewer #4), the results of our study do not change (p-value on LME model analysis = 0.04, and p-value on one-sided paired Wilcoxon rank sum test = 0.03) (see response to Reviewer #2, Row ID #9).
19	8) Page 10, line 13: The power analysis indicates use of a one-sided test, but the primary outcome analysis appears to have used a two-sided test. Please reconcile this	During the analysis, it became clear that the linear mixed effects (LME) model analysis was the most appropriate statistical approach to our primary outcome (see response to Reviewer #2, Row ID #9). Given that the construction of the LME is inherently two-sided, the primary outcome used a two-sided test. However, this did not change our results – the one-sided paired Wilcoxon signed rank

ID	Comment/Request	Response
	discrepancy.	test planned per our Research Protocol gives the same result as the LME.
20	9) Page 11, line 15: The results indicate that 25 participants was the planned sample size, but the sample size description indicates that 25 pairs were required for adequate power. Please reconcile this discrepancy.	Please see response to Editor comment, Row ID # 4. To achieve the desired effect size, significance level, and power, 25 pairs were required. In this study, a pair is made up of two nights of observation (two polysomnography studies). Inherent to the crossover design is that each participant individually makes up a pair, that is, 25 participants is equivalent to 25 pairs of polysomnography studies since each participant undergoes two nights of observation (treatment, sham-control) and is, as such, her own control (treatment-control pair), which minimizes variability and increases the power to detect changes in the primary outcome.
21	10) Page 11, Sample Characteristics: Please provide information about the length of time between sleep studies for the same participant.	Noted. In our Research Protocol, we indicated that the two study nights "*need not be consecutive nights as this may be onerous on the participants, who may have children at home*" and that we would "*accommodate the schedules and wishes of the participants*" (see Section **1.4.1.1 Methods**). Note that we also described this in our original manuscript (see **METHODS** > **Interventions**): "*Each participant underwent two overnight PSG studies (not required to be consecutive dates) between 28-37 weeks gestation.*" The length of time between studies for the same participant was a median of 1 day (IQR 1 – 3.25 days; maximum 13 days). We added the following sentence to the manuscript (see **RESULTS**): • "*The length of time between studies for the same participant (washout period) was a median of 1 day (IQR 1 – 3.25 days; maximum 13 days).*"
22	11) Page 12, Table 1: Assuming that the 3 participants who did not complete both sleep studies were excluded from analysis, the more relevant population for Table 1 is the 20 who were analyzed. The authors should also include differences between those who did and did not complete the study.	Noted. We have revised **Table 1** to only include the twenty participants who completed the study and were analyzed. Note that the two randomized groups consisting of participants who completed the study (n=10 each group) were well balanced with respect to the sample characteristics – there were no statistically significant differences in baseline characteristics between groups (α=0.05). With regard to differences in baseline characteristics between the participants who completed the the study (n=20) and those who did not (n=3), we give a narrative description of the three individuals and how they differed because statistical testing would be unreliable. The three participants who did not complete the study were younger (mean 26.7 years), had higher current BMI (mean 32.5 kg/m2), and had less self-reported overnight sleep duration at the time of the 1st polysomnography test (mean 6.8 hours); however, they were similar in ethnicity (all Caucasian), pre-pregnancy BMI (mean 26.6 kg/m2), gestational age (mean 30.6 weeks), gravidity (all G1), and sleep habits (sleep onset and waking position in the last week and when not pregnant, bed partner, snoring, pillow use). We added the following sentence to the

ID	Comment/Request	Response
		manuscript (see **RESULTS** > **Sample Characteristics**): • "*In comparing baseline demographic, obstetric, and sleep habit characteristics between the participants who completed the study (n=20) and those who did not (n=3), the groups were similar; however, participants who did not complete the study were younger (mean 26.7 years), had a higher current BMI (mean 32.5 kg/m2), and had less self-reported overnight sleep duration (mean 6.8 hours).*"
23	12) Page 16, line 26: Please report whether the primary outcome was normally distributed, since this has potential impact the appropriateness of the analysis method.	Noted. The primary outcome was not normally distributed. We added the following sentence to the manuscript (see **RESULTS** > **Primary Outcome**): • "*The primary outcome was not normally distributed.*"
24	13) Results: Please also report the length of time spent supine while sleeping by treatment.	We have added the following to the manuscript (see **RESULTS** > **Secondary Outcomes** and **Table 3**): • Length of time (minutes) supine when sleeping by treatment • Length of time left lateral when sleeping by treatment • Length of time right-lateral when sleeping by treatment Note that we had originally reported length of time supine during time in bed (TIB), where TIB is defined as the elapse time between "lights off" and "lights on" minus any time the participant was out of bed to use the washroom; however, since we are specifically interested in whether the PrenaBelt can modify position while asleep, we decided to remove the TIB-based supine time from the manuscript (**RESULTS** > **Secondary Outcomes**, **Table 3**, **DISCUSSION** > **Principal Findings**) and replace it with the time supine while asleep as requested by the reviewer.
25	14) Page 20, line 55: What is the reference for the statement that "most pregnant women continue to spend a significant amount of time supine during sleep in late pregnancy?"	This statement is based off our current study and three other studies, which we have clarified in the manuscript to read: • *"…most pregnant women continue to spend a significant amount of time supine during sleep in late pregnancy per our study and previous studies.(6–8)*" We have added the citations (n=3) of the other studies for this statement. Note that these three references were cited in other parts of our original manuscript (see **INTRODUCTION** > **Background** and **DISCUSSION** > **Strengths and Weaknesses in Relation to Other Studies**), so no additional references were added to the manuscript, they were just re-cited: • O'Brien LM, Warland J. *Typical sleep positions in pregnant women*. Early Hum Dev; 2014;90(6):315–7. • McIntyre JPR, Ingham CM, Hutchinson BL, Thompson JMD, McCowan LM, Stone PR, et al. *A description of sleep behaviour in healthy late pregnancy, and the accuracy of self-reports*. BMC Pregnancy Childbirth; 2016;16(1):115. • Warland J, Dorrian J, Gordon S, Grimmer K, Trott P, Kerr M, et al. *Accuracy of Self-Reported Sleep Position in Late Pregnancy*.

ID	Comment/Request	Response
		PLoS One; 2014;9(12):e115760.
26	15) Page 21, line 8: Please clarify whether the statement that "pregnant women's estimates of time in each sleep position...underestimate the time they spend supine" is based on the current study (which may not be generalizable due to the small sample size) or provide a reference.	This statement is based on the current study as the current study is the first to investigate pregnant women's perception of time spent sleeping supine. We have clarified this statement in the manuscript to read: • "*Also, per our study, pregnant women's estimates of time in each sleep position, while relatively accurate for lateral sleeping positions, underestimate the time they spend supine.*"
27	16) The potential for period effects in this crossover study is alluded to in the discussion of study strengths and weaknesses. However, it warrants more attention in the methods, results, and discussion given that this is a crossover design study. In this study, one could expect period effects due to a variety of factors including increased patient knowledge and comfort with the process, and increased discomfort sleeping due to being further along in pregnancy. Similarly, potential carryover effects and the length of the washout period should be explicitly addressed in the methods and results sections (and discussion if appropriate).	See response to Reviewer #4, Row ID #21. In our Research Protocol (see Section **1.4.1.1 Methods**) and original manuscript (see **METHODS** > **Interventions**), we indicated that the two study nights did not need to be consecutive in order to accommodate the participants' schedules. In our study, we did not incorporate a run-in measurement of baseline sleep habits (i.e., a non-intervention polysomnogram conducted before the two study nights) as this was not financially or logistically feasible. Also, we did not specify a set washout period. We contend that while period effects may have resulted from "increased patient knowledge and comfort with the process", they were much less likely to result from "increased discomfort sleeping due to being further along in pregnancy" given that the length of time between studies for the same participant (washout period) was a median of 1 day (IQR 1 – 3.25 days; maximum 13 days), which is negligible in the context of a 12 week long third trimester. We collected data for each patient on two different nights, so we incorporated whether it was the first or second night as a random effect in our Linear Mixed Effects (LME) model. This was to control for within-subject variation in our measured outcomes due to increased familiarity with the process (carryover). We added the following sentences: In **METHODS** > **Interventions**: • "*There was no run-in measurement of baseline sleep habits.*" • "*The study nights were not required to be consecutive dates, and we did not specify a defined washout period*." In **RESULTS**: • "*The length of time between studies for the same participant*

ID	Comment/Request	Response
		(washout period) was a median of 1 day (IQR 1 – 3.25 days; maximum 13 days)." In **DISCUSSION** > **Weaknesses**, our original manuscript acknowledges that we had no baseline data (run-in measurement). We added: • "*Period effects and carryover may have occurred due to advancing gestation and familiarization with the sleep environment/equipment; however, given the median washout period (1 day) was negligible in the context of a 12-week third trimester, period effects from the washout were, at most, minimal. Our LME model incorporated the PSG night (1st or 2nd) as a random effect to control for within-subject variation in our measured outcomes due to increased familiarity with the process*."
28	17) Consort: Diagram should provide reasons for the 3 dropouts.	Noted. We edited this image (Figure 2) to include the reasons for the three dropouts and attached the updated image to our submission.
29	n/a	We changed the description of the primary outcome from "*Proportion* of sleep time supine" to "*Percentage* of sleep time supine (%)" because this is more accurate since the outcome was reported as a percentage in our original manuscript.
30	n/a	In **INTRODUCTION** > **Background**, we changed the following sentence: • "*Recently, four studies have demonstrated an association between maternal supine sleep position in late pregnancy and the risk of late term SB (1–4) and LBW.(1) Results of a fifth study (5) are currently under review. The population attributable risk of supine sleep for SB has been reported as being between 9.4% and 37%,(1–4) suggesting…*" To: • "*Recently, five studies have demonstrated an association between maternal supine sleep position in late pregnancy and the risk of late term SB (1–5) and LBW.(1) The population attributable risk of supine sleep for SB has been reported as being between 3.7% and 37%,(1–5) suggesting…*" We also added the reference for the recently published paper (reference #5): • Heazell AEP, Li M, Budd J, Thompson JMD, Stacey T, Cronin RS, Martin B, Roberts D, Mitchell EA, McCowan LME. *Association between maternal sleep practices and late stillbirth – findings from a stillbirth case-control study*. BJOG 2018; 125:254–262.
31	n/a	Added a subsection "Patient and Public Involvement" under the **METHODS** section. Added the following statements: • "*Patients and the public were not involved in the development of the research question or outcome measures, design of the study, recruitment process, or conduct of the study. During the consent process, participants indicated whether they wished to receive a copy of their personal study results and/or a summary of the*

ID	Comment/Request	Response
		overall study results (to be shared by email)."

VERSION 2 – REVIEW

REVIEWER	Timor Faber Erasmus University Medical Centre, Rotterdam, The Netherlands
REVIEW RETURNED	12-Apr-2018

GENERAL COMMENTS	The authors have clearly incorporated most of the feedback that was given previously, and the manuscript has improved as such. However, I am still hesitant about the following: the reporting of differences in means of the primary outcome, as the distributions of these means are highly skewed. I believe a specialist statistical review will be able to provide a definite answer to whether this is the appropriate statistical method.

REVIEWER	Marie Gantz RTI International, USA
REVIEW RETURNED	06-Apr-2018

GENERAL COMMENTS	As stated in my original review, it is not acceptable for the primary outcome reported in the manuscript to differ from the primary outcome described in the protocol. If the authors wish to present a justification for an alternate definition in the manuscript they could do that, but they should make it clear what the pre-specified primary outcome was. The manuscript states that "No changes were made to the trial outcomes after trial commencement" which does not appear to be true. In addition, the authors state that a linear mixed model does not assume normality of the outcome distribution, but in fact it does. Because the the primary outcome was not normally distributed, the Wilcoxon test, as specified in the protocol, is the more appropriate choice for analysis.

VERSION 2 – AUTHOR RESPONSE

ID	Comment/Request	Response
	Thank you very much for addressing our previous comments.	You are welcome.
1	We still have concerns about the description of the study design. You have described this as a pilot study. Our instructions for authors states that "papers reporting pilot studies should explain the work's wider	We described our study as a "pilot trial" in response to the Editor's comment on our manuscript (ID bmjopen-2017-020256), Re: "*We are concerned that these findings are a little preliminary and that this might be better described as a pilot on which a further study could be planned*."

ID	Comment/Request	Response
	context and explain why the term 'pilot study' applies. The term 'pilot study' should not be applied to justify reporting a small-scale study. Justifications for a pilot study include: 1. trialing a new procedure intended for use in a larger programme of research 2. establishing power calculations required for a full-scale study 3. establishing how many patients and/or healthcare professionals can be recruited 4. evaluating the financial, technical, administrative or logistic feasibility of a full-scale study, including issues of data collection, protocol adherence, and questionnaire design. The sample/patient size should still be justified. The article should explain the impact that the pilot study had on decisions regarding future research." Please also see the following paper about pilot studies: https://onlinelibrary.wiley.com /doi/full/10.1111/j..2002.384.doc.x	Thank-you for pointing out BMJ's instructions for authors regarding the pilot-study-justifications #1 through #4 for pilot studies. Thank-you for also pointing us to the paper about pilot studies by Lancaster, et al. in 2001. We have reviewed this paper and its recommendations in full. Further, we have reviewed Arain, et al. (BMC Medical Research Methodology 2010, 10:67), in which the authors repeated Lancaster, et al.'s review to provide a more updated analysis of current practice and editorial policy surrounding pilot and feasibility studies. Most importantly, we also reviewed Eldridge, et al. (PLoS ONE 2016, 11(3): e0150205), on which Lancaster is a co-author and provides the most up to date definition and conceptual framework for these studies. In fact, we feel it is appropriate to summarize here Eldridge, et al.'s proposal that "feasibility" is an overarching concept within which three distinct types of study can be defined: 1. **Randomised pilot studies** "*are those studies in which the future RCT, or parts of it, including the randomisation of participants, is conducted on a smaller scale (piloted) to see if it can be done*." 2. **Non-randomised pilot studies** "*are similar to randomised pilot studies; they are studies in which all or part of the intervention to be evaluated and other processes to be undertaken in a future trial is/are carried out (piloted) but without randomisation of participants.*" 3. **Feasibility studies that are not pilot studies** "*are those in which investigators attempt to answer a question about whether some element of the future trial can be done but do not implement the intervention to be evaluated or other processes to be undertaken in a future trial, though they may be addressing intervention development in some way… Within the framework, these studies can be called feasibility studies but cannot be called pilot studies since no part of the future randomised controlled trial is being conducted on a smaller scale.*" Eldridge, et al. "*suggest that researchers view feasibility as an overarching concept, with all studies done in preparation for a main study open to being called feasibility studies, and with pilot studies as a subset of feasibility studies. All such studies should be labelled '**pilot**' and/or '**feasibility**' as appropriate, preferably in the title of a report, but if not certainly in the abstract.*" Given our review of the BMJ Instructions for Authors and the three landmark papers cited above, we think that our study should be labelled as a "randomized pilot study" since it included conducting parts of a future RCT on a smaller scale to

ID	Comment/Request	Response

	In light of the above information, we would be grateful if you could elaborate further on why this study is described as a pilot trial.	see if it could be done.
		
		Reviewing the *original* intent of our study, per our Research Protocol, note that it meets the above definition of "randomized pilot trial" and the BMJ Instructions for Authors pilot study justifications #1, #2, and #4:
		
		• #1: trialing a new procedure intended for use in a larger programme of research
		  ○ There has never been a positional therapy sleep study performed on pregnant women. Our original proposal to our funders was to complete two PrenaBelt studies sequentially: first, our PrenaBelt study in Halifax (Canada, proposed n=30, a sham-controlled, double-blind, cross-over, randomized trial in a sleep lab over two nights - bmjopen-2017-020256.R1), which would be subsequently followed by our PrenaBelt study in Accra (Ghana, proposed n=200, a sham-controlled, double-blind, randomized trial in participants' own homes over 12 weeks of the third trimester - bmjopen-2018-022981). The study in Halifax was originally to inform the study in Accra; however, due to unforeseen events that threatened our ability to meet our funder's timeline constraints, we were forced to complete the two studies simultaneously.
		• #2: establishing power calculations required for a full-scale study
		  ○ From our Research Protocol, **Section 1.4.1.3 Outcomes**: *The following data will be collected from each participant across each intervention and serve as pilot data to inform effect size calculations for future research…*
		  ○ From our Research Protocol, **Section 1.4.1.5 Sample Size**: *…this study will be used to generate preliminary data to be used in effect size calculations for future clinical trials…*
		  ○ Because positional therapy in pregnancy is a novel area of study, we did not define the specifics of any future full-scale study at the time of writing the Research Protocol for the current study but simply stated that the design of future studies may look to the current study for effect size calculations.
		  ○ There are only about a dozen researchers around the world studying positional therapy in pregnancy. We are aware of at least one, multicenter, international trial currently being planned by our collaborators in the area of positional therapy in pregnancy research. The corresponding author has provided information for effect size calculations to the lead researcher planning this upcoming trial.
		• #4: evaluating the financial, technical, administrative or logistic feasibility of a full-scale study, including issues of data collection, protocol adherence, and questionnaire design
		  ○ From Research Protocol, **Section 1.3 Feasibility Study Objectives**: *Obtain PrenaBelt user experience feedback to evaluate PrenaBelt feasibility and acceptability and optimize PrenaBelt design for future research*.

ID	Comment/Request	Response
	It should also be clearer why your study should be considered as a	○ The feedback we collected from our participants regarding the PrenaBelt was technical in nature and primarily for the purpose of improving the PrenaBelt design for use in a future full-scale study, but again, we did not define the specifics of this future study in our Research Protocol. We have edited the **Title**, **Abstract**, **INTRODUCTION**, **METHODS**, and **DISCUSSION** sections to state that the current study is a "randomized pilot trial" per Eldridge, et al.'s recommendation and have provided clarification for why it should be considered a randomized pilot trial in the **INTRODUCTION** section: • Added sentence and cited (Eldridge et al., 2016): *This study is a randomized pilot trial because it trialed a new intervention on a smaller scale (pilot) to evaluate it for use in a full-scale randomized controlled trial.(33)*

ID	Comment/Request	Response
	pilot trial in the introduction section.	
2	The reporting of the study's outcomes still needs improving. For example, your secondary outcome is not clearly described in the methods >> outcomes section. You have also not made it clear that you looked at other outcomes in this section. Please also see reviewer [4]'s comment below about the primary outcome. Any discrepancies between the protocol and the manuscript should be identified and explained in your paper. This should include the reasons for not reporting the results of some of the other outcomes registered in the clinical trials registry.	We have re-written the **METHODS** > **Outcomes** section to specifically and clearly describe the primary outcome (both pre-specified and final and reasons for discrepancy – see response to Reviewer #4, Row ID #4) and secondary outcomes so that they correspond to the exact order that they are presented in the **RESULTS** section. As for the equipment used to measure these outcomes, which was listed in the **METHODS** > **Outcomes** section in our original manuscript (bmj-open-2017-020256.R1), we have relocated this to the **METHODS** > **Intervention** section. Noted. See response to Reviewer #4, Row ID #4 (below). In our trial registry, which was based off our research protocol, we registered 44 outcomes. In our manuscript, we reported on 30 outcomes. The 14 outcomes not reported in our manuscript and reasons for not reporting these were described in our first response to the BMJ Open Editor (Row ID #2). We have edited our manuscript to identify and explain these 14 outcomes and the discrepancy as follows: • In **METHODS** > **Outcomes**, we added a paragraph: ○ "*Per our research protocol and trial registry, supine and non-supine apnea-hypopnea index (AHI), min/mean/max SpO2 while awake, and min/mean/max maternal heart rate during wake, REM, and non-REM states, were also specified as secondary outcomes; however, we were unable to report the supine and non-supine AHI and heart rate data due a software configuration issue nor the awake SpO2 values due to data artifact.*"
3	The authors have clearly incorporated most of the feedback that was given previously, and the manuscript has improved as such. However, I am still hesitant about the following: the reporting of differences in means of the primary outcome, as the distributions of these means are highly skewed. I	Thank-you. See response to Reviewer #4, Row ID #5 (below). We have edited the manuscript to <u>delete</u> the reporting of differences in

ID	Comment/Request	Response
	believe a specialist statistical review will be able to provide a definite answer to whether this is the appropriate statistical method.	means of the primary outcome given the non-normal distribution of these means. Instead, we have reported the medians and IQR's of the primary outcome and used a non-parametric (one-sided, paired, Wilcoxon signed rank) test for differences, which yields a "pseudomedian", which is the median of the pairwise differences in the primary outcome between the two groups (treatment and sham), and a 95% confidence interval for the pseudomedian. We have also updated **Supplementary File 2: Code and Output – PSG Analysis** accordingly. A specialist statistical reviewer (Kara Matheson, Senior Biostatistician, Research Methods Unit, Nova Scotia Health Authority) suggested that our statistician (MB) could apply transformations or non-linear methods to address your concern here; however, we ultimately decided to follow Reviewer #4's request to adhere to our pre-specified analysis for non-normal outcomes per our research protocol because we felt this was a more appropriate approach (than applying transformations or non-linear methods) in the context of a randomized pilot trial and would also address your concern here.
4	As stated in my original review, it is not acceptable for the primary outcome reported in the manuscript to differ from the primary outcome described in the protocol. If the authors wish to present a justification for an alternate definition in the manuscript they could do that, but they should make it clear what the pre-specified primary outcome was. The manuscript states that "No changes were made to the trial outcomes after trial commencement" which does not appear to be true.	We reiterate our previous response for context: "*At the time we wrote our research protocol… we included right-lateral sleep as part of the primary outcome but made provision for excluding it based on our Research Ethics Board's (IWK REB) recommendation and participant comfort.…to avoid confusion for our readers, we contend that the primary outcome should solely be "Percent of sleep time supine (%)", which is consistent with the Research Protocol given the provision that was made and deemed acceptable by the IWK REB at the time of approval*." As stated in our manuscript, the PrenaBelt used in this study was configured to only cause supine pressure points and could not cause right-sided pressure points. We have made it clear what the pre-specified primary outcomes were and presented a justification for narrowing the primary outcome in the manuscript. We have <u>updated</u> the **METHODS** > **Outcomes** section to state: • *The pre-specified primary outcomes per protocol were the percent of time spent in the supine and right-lateral positions (as we originally intended to minimize both) and the PrenaBelt user feedback questionnaire; however, for reporting clarity and participant comfort, the PrenaBelt was configured to provide pressure points only when the user was*

ID	Comment/Request	Response
		supine, the primary outcome was further specified to be the percentage of time spent supine during sleep only, and the other pre-specified primary outcomes were instead reported as secondary outcomes. The primary outcome was documented continuously by an RA via video feed. We have removed the sentence from **METHODS** > **Outcomes**: • "*No changes were made to the trial outcomes after trial commencement*"

5	In addition, the authors state that a linear mixed model does not assume normality of the outcome distribution, but in fact it does.
	
	Because the primary outcome was not normally distributed, the Wilcoxon test, as specified in the protocol, is the more appropriate choice for analysis.
		Agreed. This was an error on our part.
		
		We have updated the manuscript to remove the linear mixed effects model and, given the non-normal distribution of the primary outcome, we have replaced it with the Wilcoxon test (non-parametric) as pre-specified in our research protocol:
		
		• **METHODS** > **Statistical Methods**
		  ○ Deleted paragraph beginning with, "*In the primary analysis, we specified a linear mixed-effects model…*"
		• **RESULTS** > **Primary Outcome**
		  ○ Deleted sentence beginning with, "*The linear mixed-effects model estimate…*" and replaced it with: "*On a one-sided paired Wilcoxon signed rank test, the median of the pairwise differences in percentage of sleep time supine between the sham night versus the PrenaBelt night was significantly greater than zero (pseudomedian=5.8, p=0.03).*"
		  ○ Updated Table 2 per above.
		• **RESULTS** > **Secondary Outcomes**
		  ○ Updated Table 3 to remove the linear mixed-effects model and replace it with the Wilcoxon test (this pertained to the row containing "Supine sleep (minutes)" and the descriptive text below Table 3).
		• **DISCUSSION** > **Weaknesses**
		  ○ Deleted sentence beginning with, "*Our LME model incorporated the PSG night...*"
		• We have updated **Supplementary File 2: Code and Output – PSG Analysis** accordingly.

6	n/a	Added Dr. Jerry Coleman to the author list and updated **OTHER INFORMATION** > **Contributors** and **OTHER INFORMATION** > **Competing Interests** accordingly. The corresponding author, Allan Kember, requested this change to authorship on 19MAR2018 and received a reply from the BMJ Editorial Production Assistant on 26MAR2018 with instructions. All eight co-authors approved the requested change by responding to the

ID	Comment/Request	Response
		Editor by 04APR2018.
7	n/a	Updated contact information of corresponding author as he is relocating in June 2018.
8	n/a	Updated **OTHER INFORMATION** > **Competing Interests** section to be consistent with the same section in our other manuscript submitted to BMJ Open (manuscript ID# bmjopen-2018-022981).

VERSION 3 – REVIEW

REVIEWER	Marie Gantz RTI International
REVIEW RETURNED	08-Jun-2018

GENERAL COMMENTS	Thank you for addressing concerns raised in the previous reviews. For clarity and completeness, I recommend including in the statistical methods that one-sided tests were used.