Supplementary File 3: Additional Discussion

LEFT SHIFT IN CBWC BETWEEN V6.7.8.1 AND V8.0.6.2 GROW CALCULATORS

We acknowledge that the median CBWC dropped (left shift) from 43% (v6.7.8.1 calculator) to 42% (v8.0.6.2 calculator) in the treatment group and from 31% (v6.7.8.1 calculator) to 28% (v8.0.6.2 calculator) in the sham group. The “Ghanaian” ethnicity coefficient was not yet available in the v6.7.8.1 calculator, so the authors used the regional “West African” coefficient for the original GPT analysis, as did we when we replicated it (see Table 1 in our main text). The “West African” coefficient in the v6.7.8.1 calculator was derived from a 10 year database of routine, scan-dated, pregnancies booked and delivered at a single unit (Queen’s Medical Centre, Nottingham, UK).(1) While these data represent West Africans living in the UK, analyses have shown that differences between 1st and 2nd generation migrants are negligible.(1) However, the new centile calculators,(2) including the v8.0.6.2 calculator,(3) allow calculation of CBWC and use an ultrasound-derived estimated fetal weight, for 120 ethnic, country of origin, and region coefficients.(4) The v8.0.6.2 calculator is based on data from over 4 million pregnancies submitted from 33 countries.(4) The “Ghanaian” ethnicity coefficient in the v8.0.6.2 calculator is, therefore, directly based on datasets from Ghana, not West Africans living in the UK as with the v6.7.8.1 calculator. We submit that this key difference between the old (v6.7.8.1) and new (v8.0.6.2) calculators likely accounts for the left shift in the CBWC results. Furthermore, as GROW (Gestation-Related Optimal Weight) sets an “optimal” standard, overall one would expect the CBWC distribution to be left shifted and have more SGA than LGA cases, as it is more common for the growth potential to be not reached than to be exceeded. For reasons cited above, we

also submit that our results with the new (v8.0.6.2) calculator are more representative of the true population.

COMPARISON OF FREQUENTIST AND BAYESIAN PARADIGMS

Null Findings and Sensitivity in the Frequentist Framework

Note that in the course of experimentation, the null hypothesis is never proved or established but is possibly disproved.(5) A type I statistical error, also known as a false positive, occurs when investigators state that a treatment effect exists when it, in fact, does not.(6) A type II statistical error, also known as a false negative, occurs when investigators state that a treatment effect does not exist when, in fact, it does. When a treatment is relatively safe, as is the case with positional therapy (PT), the consequences of committing a type I error are minimal. In contrast, however, the consequences of committing a type II error can be substantial because it may result in patients being denied a treatment with beneficial effects, and may also suggest that further research is not required.(7) Tragically, there are several examples of the latter case occurring in the literature, including fibrinolytics.(8,9)

Whenever null findings are reported, it is important to complete a sensitivity analysis. The 95% confidence interval (95% CI) of the sample mean difference, which was -0.3% to 18.2% in our study, tells us the sensitivity of the experiment directly: if it includes both the value of the null hypothesis (0%) and other interesting effect sizes, e.g., the alternate hypothesis, then the experiment was not sensitive enough to draw definitive conclusions.(10) This is inherent in the true meaning of the 95% CI, which is often misinterpreted by readers of scientific literature as a de facto significance test by

examining whether its endpoints overlap the null value.(11) The 95% CI is computed by finding the set of all values of the dependent variable (customised birthweight centile [CBWC] in our study) that are non-significantly different from the sample mean difference (8.4% in our study) at the 5% level. That is, all the points in our 95% CI (-0.3% to 18.2%) are non-significantly different from the sample mean difference (8.4%), which tells us that the GPT data are consistent with PT both having no effect (0%) and with PT having an effect as large as an 18.2% gain in CBWC. Outside the interval of -0.3% to 18.2%, out to infinity in both directions, all the values are significantly different at the 5% level from the sample mean difference of 8.4% and can be ruled out as possible population values. However, all the points within the 95% CI of the mean difference in CBWC per the GROW standard (version 8.0.6.2) with PT versus sham-PT (-0.3% to 18.2%) cannot be ruled out as population values. As discussed in **Supplementary File 2**, while there is no consensus on the minimum clinically important difference (MCID) in CBWC, our 95% CI contains effect sizes that most clinicians would agree are important clinically such as 5%, 10%, and 15%. For example, a 5% gain in CBWC would boost a growth restricted foetus (<3rd centile by definition) into the 3rd to <10th centile range, halving its stillbirth risk, and a 10% gain in CBWC would boost it into the normal range (10th to 90th centile), reducing its stillbirth risk by five fold.(12)

Approach to Data and Hypotheses

Approaching the data from a different analytical paradigm (Bayesian), however, indicates that it is highly probably or almost certain that nightly maternal PT, compared to sham-PT, during sleep throughout the third trimester confers a significant benefit to foetal growth – even for the sceptic, benefit (>0% gain in CBWC) was highly probable

(0.81). This apparent discrepancy in the results under the frequentist paradigm compared to the Bayesian paradigm is not so much a discrepancy in the results as it is a discrepancy in the way that clinicians think and in the way these two paradigms approach data and hypotheses.

The frequentist asks, "*Does my data fit my hypothesis?*", whereas the Bayesian asks, "*Does my hypothesis fit my data?*". The p-value (frequentist paradigm) is the probability that if the experiment were repeated an indefinite number of times, we would observe results as extreme or more extreme than the results we observed assuming that the null hypothesis is true (i.e., "*Does my data fit my hypothesis?*").(10) As such, the p-value is a long-run frequency; however, clinicians think in terms of conditional probabilities (i.e., "*Does my hypothesis fit my data?*", or in other words, "*What is the probability of my hypothesis being true given my data?*"), not long-run frequencies.(13)

When a clinician encounters a patient, they perform a history and physical exam (data) and estimate a pre-test probability that the patient has a given disease (hypothesis). The clinician orders further investigations of the patient, and based on these results (new data), the post-test probability of disease (hypothesis) is revised, and so on. At no point in this process is the clinician thinking about the patient, the data, and the hypothesis in terms of long-run repeated experiments with a p-value.(6) Furthermore, when it comes to treating a disease with an intervention, clinicians are more interested in knowing the probability that the intervention is effective given the available data (Bayesian thinking) and less interested in the long-run frequency of observing data as extreme or more extreme than that observed previously assuming that the intervention is ineffective (frequentist thinking).

92 **Justification of Reanalysis of a "Negative Trial" with Bayesian Methods**

93 A Bayesian reanalysis of a "negative" trial under the frequentist paradigm does not
94 create positive results nor should it be employed as a statistical alternative in an effort to
95 demonstrate a treatment effect. To demonstrate, it is important to note that if PT had no
96 effect, the posterior probability of PT being better than sham-PT in a Bayesian analysis
97 would be 0.50 on average. That is, a truly futile intervention (not helpful nor harmful)
98 would result in 50% of the area under the posterior probability curve (lighter shade) in
99 **Figure 2** (see **Results** section in main text) being to the left of the 0% threshold and
100 50% being to the right of the 0% threshold. This was not the case seen in our analysis
101 where the bulk of the area under the posterior probability curve – for a menu of priors
102 reflecting varying levels of enthusiasm and scepticism – is located to the right of the 0%
103 threshold, which indicates a treatment benefit. As such, we made a deliberate choice to
104 use Bayesian methods to reanalyze the GPT data because frequentist methods had
105 previously provided an incomplete summary of the results.

106 **LIMITATIONS**

107 **Exclusion of Birth Weight From Bayesian Analysis**

108 The original GPT publication had two primary outcomes: birth weight and CBWC.(14)
109 We did not complete a Baysian analysis of the raw birth weight and, instead, chose the
110 CBWC for analysis for reasons described in the **Limitations** section of the main text.
111 That said, given the p-value (0.14), difference (gain of 110 grams), and 95% CI for the
112 treatment effect of PT on birth weight (−38 to 258 grams) in the original GPT

publication,(14) a Bayesian reanalysis of the GPT birth weights would likely show a beneficial treatment effect of PT on birth weight as we saw with CBWC, especially in light of the birth weight findings in Anderson et al.'s meta-analysis.(15) This meta-analysis reported that for n=57 participants whose going-to-sleep position in the third trimester (after 28 weeks) was supine, their infant's mean (standard deviation) birth weight centile was 3410 (112) grams, and for n=1703 participants whose going-to-sleep position in the third trimester was non-supine, their infant's mean (standard deviation) birth weight centile was 3554 (98) grams. Comparing these two groups, the adjusted mean difference in birth weight was a gain of 144 grams (95%CI: 36 to 253 grams; p-value 0.009) with non-supine going-to-sleep position.

Data-Derived Prior from Meta-analysis

We used data from a recent individual participant data meta-analysis of sleeping position in the third trimester (15) to inform our strongly enthusiastic (SE) prior (see the **Methods** section in main text) for the Bayesian analysis. Following our methodology, based on this meta-analysis, we would approach the GPT with a belief that nightly use of PT from 28 weeks to birth increases CBWC by 9% on average (probability 0.50) compared to sham-PT. This prior distribution corresponds to a probability of 0.99, 0.85, and 0.40 that PT improves CBWC by 0%, 5%, and 10% or more compared to sham-PT. This meta-analysis is a good source of information on which to base our prior belief vis-a-vis the impact of sleeping position on CBWC for at least two reasons: first, it analyses the CBWC using the customised growth standard that we used (GROW), and second, its inclusion criteria stipulated that participants must be at 28 weeks gestation or greater

at the time of interview about their sleeping position, which was the same eligibility criteria employed for participation in the GPT.

However, one major limitation is that the four studies included in the metaanalysis were observational (case-control) studies and, as such, did not use PT and, rather, analysed CBCW in light of the going-to-sleep position. While the going-to-sleep position does not tell the whole story about what happens to sleeping position after sleep onset, a recent study by Wilson et al. gives helpful and relevant insight: in the third trimester, the going-to-sleep position is the dominant sleeping position overnight in the majority (54%) of pregnancies, and the supine going-to-sleep position results in significantly more time spent sleeping supine overnight compared to a non-supine going-to sleep position (48% vs. 22.6%, $p<0.001$).(16) While these findings do not completely assuage this limitation of informing our SE prior from Anderson et al.'s meta-analysis, it may attenuate it until data are available from other interventional trials of PT in pregnancy. Note that as of the writing of the main text of this manuscript, the GPT is the only interventional trial of PT during sleep in pregnancy with foetal growth as an outcome.

Another limitation is that the participant samples of the four studies included in the meta-analysis were drawn from populations in New Zealand, Australia, and the United Kingdom, which are different from the participant sample in the GPT (Ghana). While the GROW CBCW, which the meta-analysis used, accounts for maternal ethnicity and other population-related factors (e.g., maternal height and weight), there may be other population-related factors that affect foetal growth (e.g., socioeconomic status). As such, sleeping position may not affect foetal growth in the same way in different populations; however, one study from the same setting and centre as the GPT,

158 indicates the contrary, at least in the Ghanaian population, and found that the newborns
159 of participants who reported supine sleep during pregnancy were at a five-fold
160 increased risk of low birth weight.(17)

161 **REFERENCES**

162 1. Gardosi J, Francis A. A Customised Weight Centile Calculator - Gestation Related
163 Optimal Weight [Internet]. Gestation Network; 2016. Available from:
164 www.gestation.net
165 2. Gardosi J. GROW 1.5 Documentation December 2020 [Internet]. Gestation Network
166 and Perinatal Institute; 2020 [cited 2024 Feb 10]. Available from:
167 https://www.gestation.net/GROW_documentation.pdf
168 3. Gardosi J, Hugh O, Francis A. Gestation Related Optimal Weight Bulk Centile
169 Calculator [Internet]. Gestation Network; 2022. Available from: www.gestation.net
170 4. Gardosi J. GROW v8.0.6.2 (2022) – User Guide [Internet]. Gestation Network and
171 Perinatal Institute; 2022 [cited 2024 Feb 10]. Available from:
172 https://www.gestation.net/BCC_User_Guide_v8.0.6.2.pdf
173 5. Fisher RA. The design of experiments. Oxford, England: Oliver & Boyd; 1935. xi, 251
174 p. (The design of experiments).
175 6. Goodman SN. Toward evidence-based medical statistics. 1: The P value fallacy. Ann
176 Intern Med. 1999 Jun 15;130(12):995–1004.
177 7. Alderson P, Chalmers I. Survey of claims of no effect in abstracts of Cochrane
178 reviews. BMJ. 2003 Mar 1;326(7387):475.
179 8. Altman DG, Bland JM. Statistics notes: Absence of evidence is not evidence of
180 absence. BMJ. 1995 Aug 19;311(7003):485.
181 9. Yusuf S, Collins R, Peto R, Furberg C, Stampfer MJ, Goldhaber SZ, et al.
182 Intravenous and intracoronary fibrinolytic therapy in acute myocardial infarction:
183 overview of results on mortality, reinfarction and side-effects from 33 randomized
184 controlled trials. Eur Heart J. 1985 Jul;6(7):556–85.
185 10. Dienes Z. Understanding Psychology as a Science. New York, USA: Palgrave
186 Macmillan; 2008.
187 11. Johnson SR, Feldman BM, Pope JE, Tomlinson GA. Shifting our thinking about
188 uncommon disease trials: the case of methotrexate in scleroderma. J Rheumatol.
189 2009 Feb;36(2):323–9.
190 12. Agarwal U, Hugh O, Gardosi J. Prospective risk of stillbirth according to fetal size
191 at term. J Perinat Med. 2022 Jul 1;50(6):748–52.
192 13. Burton PR, Gurrin LC, Campbell MJ. Clinical significance not statistical
193 significance: a simple Bayesian alternative to p values. J Epidemiol Community
194 Health. 1998 May;52(5):318–23.
195 14. Coleman J, Okere M, Seffah J, Kember A, O'Brien LM, Borazjani A, et al. The
196 Ghana PrenaBelt trial: a double-blind, sham-controlled, randomised clinical trial to
197 evaluate the effect of maternal positional therapy during third-trimester sleep on birth
198 weight. BMJ Open. 2019 Apr 1;9(4):e022981.

199 15. Anderson NH, Gordon A, Li M, Cronin RS, Thompson JMD, Raynes-Greenow
200 CH, et al. Association of Supine Going-to-Sleep Position in Late Pregnancy With
201 Reduced Birth Weight. JAMA Netw Open. 2019 Oct 2;2(10):e1912614.
202 16. Wilson DL, Fung AM, Pell G, Skrzypek H, Barnes M, Bourjeily G, et al.
203 Polysomnographic analysis of maternal sleep position and its relationship to
204 pregnancy complications and sleep-disordered breathing. Sleep. 2022 Apr
205 11;45(4):zsac032.
206 17. Owusu JT, Anderson FJ, Coleman J, Oppong S, Seffah JD, Aikins A, et al.
207 Association of maternal sleep practices with pre-eclampsia, low birth weight, and
208 stillbirth among Ghanaian women. Int J Gynaecol Obstet Off Organ Int Fed Gynaecol
209 Obstet. 2013 Jun;121(3):261–5.