Correlation between neonatal outcomes of twins depends on the outcome: secondary analysis of twelve randomised controlled trials

Lisa N Yelland[a,b], Ewoud Schuit[c], Javier Zamora[d,e], Philippa F Middleton[a,f], Arianne C Lim[g], Anwar H Nassar[h], Line Rode[i,j], Vicente Serra[k], Elizabeth A Thom[l], Christophe Vayssière[m,n], Ben WJ Mol[o], Simon Gates[p]

[a]South Australian Health and Medical Research Institute, Adelaide, Australia

[b]School of Public Health, The University of Adelaide, Adelaide, Australia

[c]Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, The Netherlands

[d]Women's Health Research Unit, Barts and the London School of Medicine and Dentistry, Queen Mary University of London, London, United Kingdom

[e]CIBER Epidemiology and Public Health and IRYCIS, Madrid, Spain

[f]Robinson Research Institute, School of Medicine, The University of Adelaide, Adelaide, Australia

[g]Department of Obstetrics & Gynecology and GROW-School for Oncology and Developmental Biology, Maastricht University Medical Centre, Maastricht, The Netherlands

[h]Department of Obstetrics and Gynecology, American University of Beirut Medical Center, Beirut, Lebanon

[i]Department of Obstetrics, Rigshospitalet, Copenhagen University Hospital, Denmark

[j]Department of Clinical Biochemistry, Herlev and Gentofte Hospital, Copenhagen University Hospital, Denmark

[k]Maternal-Fetal Medicine Unit, Valencia Infertility Institute, University of Valencia, Valencia, Spain

[l]George Washington University Biostatistics Centre, Washington, DC, United States of America

[m]Department of Obstetrics and Gynecology, Paule de Viguier Hospital, CHU Toulouse, Toulouse, France

[n]UMR 1027 INSERM, University of Paul Sabatier Toulouse III, Toulouse, France

[o]Department of Obstetrics and Gynaecology, Monash Medical Centre, Monash Health and Monash University, Victoria, Australia

[p]Warwick Clinical Trials Unit, University of Warwick, Coventry, United Kingdom

Author for Correspondence: Dr Lisa Yelland, School of Public Health, Mail Drop DX650 511, The University of Adelaide, SA 5005, Australia, Email: lisa.yelland@adelaide.edu.au, Phone: +61 8 8313 3215.

Running Title: Correlation between neonatal outcomes of twins

Word Count: 3036

Abstract

Objective: To estimate the magnitude of the correlation between neonatal outcomes of twins and demonstrate how this information can be used in the design of randomised controlled trials (RCTs) in women with twin pregnancies.

Design: Secondary analysis of data from 12 RCTs.

Setting: Obstetric care in multiple countries, 2004-2012.

Population or Sample: 4504 twin pairs born to women who participated in RCTs to assess treatments given during pregnancy.

Methods: Intraclass correlation coefficients (ICCs) were estimated using log binomial and linear models.

Main Outcome Measures: Perinatal death, respiratory distress syndrome, bronchopulmonary dysplasia, intraventricular haemorrhage, necrotising enterocolitis, sepsis, neonatal intensive care unit admission, birthweight, low birthweight and two composite measures of adverse neonatal outcome.

Results: ICCs for the composite measures of adverse neonatal outcome were all above 0.5, indicating moderate to strong correlation between adverse outcomes of twins. For individual neonatal outcomes, median ICCs across trials ranged from 0.13 to 0.79 depending on the outcome. An example illustrates how ICCs can be used in sample size calculations for RCTs in women with twin pregnancies.

Conclusions: The correlation between neonatal outcomes of twins varies considerably between outcomes and may be lower than expected. Our ICC estimates can be used for designing and analysing RCTs that recruit women with twin pregnancies and performing meta-analyses that

include such RCTs. Researchers are encouraged to report ICCs for neonatal outcomes in twins in their own RCTs.

Funding: Australian National Health and Medical Research Council (ID 1052388).

Keywords: Sample size, power, Bayesian analysis, meta-analysis, twins, intraclass correlation coefficient

Tweetable Abstract

Correlation between neonatal outcomes of twins depends on the outcome and may be lower than expected

INTRODUCTION

Twin births and their associated complications are on the rise. In high income countries, twin births now account for around 2-4% of all births due to increasing use of assisted reproductive technologies and advancing maternal age.[1] Compared with singleton pregnancies, twins have a higher risk of adverse neonatal outcomes including preterm birth, respiratory distress syndrome, low birthweight and mortality.[2, 3] Antenatal interventions intended to improve neonatal outcomes, such as prophylactic progesterone treatment, have been studied specifically in women with twin pregnancies but with limited success.[4-9] Further randomised controlled trials (RCTs) evaluating promising interventions in this high risk population are needed.

Designing and analysing RCTs in women with twin pregnancies is challenging. Twins born to the same mother are expected to have similar or correlated outcomes due to the shared fetal and neonatal environment and common genetic material. [10, 11] As a result, infants born from the same twin pregnancy cannot be viewed as two independent trial participants and this has implications for the trial design and analysis. In particular, the correlation between outcomes of twins should be taken into account in the sample size calculations to maintain the desired power,[12] and in the analysis to avoid producing results that are over-precise.[13] The higher the correlation, the larger the impact twins have on the sample size and analysis.

An accurate estimate of the correlation between twins is important, as this is likely to vary across different outcomes and populations. Higher correlation is expected for certain outcomes, such as gestational age at birth, where the twin-to-twin delivery interval rarely exceeds one day. Higher

correlation is also expected in certain populations, such as monochorionic twin pregnancies, where twins share both their genetics and placenta. An estimate of the relevant correlation from an external source is often required. Since the correlation between neonatal outcomes of twins is rarely reported in trial publications,[14] appropriately designing and analysing RCTs in women with twin pregnancies can be difficult and published estimates are needed.

The purpose of this study is to estimate the magnitude of the correlation between neonatal outcomes of twins for commonly reported outcomes, both overall and by chorionicity. We demonstrate how this information can be used in sample size calculations for RCTs in women with twin pregnancies, as this is likely to be their most common use, and discuss other potential uses in Bayesian analyses and meta-analyses.

METHODS

Datasets

Twelve datasets including a total of 4504 twin pairs were used to estimate ICCs, as summarised in Appendix Tables S1 and S2. The datasets were from a convenience sample of RCTs chosen based on the availability of individual participant data for twins with adverse neonatal outcomes defined in a standardised manner as part of previous studies. The principal investigators of all RCTs were contacted and provided permission to use the data for the current study. The first dataset comes from a multicentre, open-label RCT assessing the effectiveness of a cervical pessary compared to no intervention for preventing poor perinatal outcomes.[15] The trial recruited 813 women with a

multiple pregnancy between 12 and 20 weeks' gestation, of whom 795 had a twin pregnancy (23% monochorionic, 77% dichorionic) and were part of this study. Exclusion criteria were known serious congenital defects, fetal death, twin-to-twin transfusion syndrome and known placenta previa. Women assigned to the cervical pessary group had a pessary inserted between 16 and 20 weeks' gestation and removed in the 36th week of gestation, while women in the control group received standard antenatal care. Approximately 55% of women delivered preterm (<37 weeks' gestation).

The remaining datasets come from 11 RCTs included in an individual participant data meta-analysis designed to investigate the effects of progestogens in women with a twin pregnancy.[16] Trials were eligible for inclusion if they compared the effect of vaginally administered progesterone or intramuscular 17-hydroxyprogesterone caproate (17Pc) versus placebo or non-intervention in the second or third trimester in women with a twin pregnancy on either preterm birth or adverse perinatal outcome. Thirteen trials met the inclusion criteria and contributed individual participant data to the meta-analysis, however, only the 11 trials that included a minimum of 40 women with a twin pregnancy were included in this study.[4-9, 17-21] Inclusion/exclusion criteria and treatment regimens varied between these trials (Appendix Table S1). The study size ranged from 67 to 677 twin pairs, with trials either including both monochorionic and dichorionic twin pregnancies,[4, 5, 7, 17, 18, 21] dichorionic twin pregnancies only[6, 8, 19] or not recording chorionicity[9, 20] (Appendix Table S2). Preterm birth rates (<37 weeks' gestation) ranged from 50-79%.

Neonatal Outcomes

For each trial, the following 12 neonatal outcomes were defined, where possible: perinatal death (intrauterine fetal death at any gestational age or neonatal death before hospital discharge); respiratory distress syndrome (RDS) requiring oxygen for at least 24 hours; bronchopulmonary dysplasia (BPD); intraventricular haemorrhage (IVH) grade III or IV; necrotising enterocolitis (NEC) grade II or higher; culture-proven sepsis; admission to the neonatal intensive care unit (NICU); birthweight; low birthweight (<2500g and <1500g) and two composite measures of adverse neonatal outcome, as defined in a previous study.[16] The first composite outcome included perinatal death, RDS, BPD, IVH, NEC and sepsis, while the second included perinatal death, RDS, IVH and NEC.

Statistical Methods

The magnitude of the correlation between neonatal outcomes of twins was measured using the intraclass correlation coefficient (ICC). An ICC of 0 indicates that neonatal outcomes of twins are completely independent and the ICC approaches 1 for neonatal outcomes typically experienced by either both or neither members of a twin pair. The data were analysed using log binomial models for binary outcomes and linear models for continuous outcomes. Adjustment was made for treatment group, since ICCs calculated ignoring potential treatment effects may be biased,[22] and a single ICC was estimated for both treatment groups combined. Clustering due to twins was taken into account using generalised estimating equations (GEEs), as this is the most common analysis approach used to account for twins in RCTs.[14, 23] ICCs were estimated by the correlation parameter for the exchangeable working correlation structure; more complex correlation structures reduce to

an exchangeable correlation structure when the cluster size is two. As a sensitivity analysis, ICCs were also estimated from linear mixed effects models with a random mother effect. Confidence intervals (CIs) for ICCs were obtained via bootstrapping using the bias corrected and accelerated method[24] with 2000 bootstrap samples and resampling of clusters (mothers), rather than individuals (infants). Each trial was analysed separately, both overall and by chorionicity where available. No analysis was performed for individual outcomes in trials where there were less than 40 sets of twins with available data for the outcome, or less than 10 cases of a binary outcome, as the ICC estimates were considered too unreliable and GEEs are known to produce biased residuals when the number of clusters is small.[25, 26] ICCs and 95% confidence intervals are presented by trial, along with the prevalence for binary outcomes and the mean and standard deviation (SD) for continuous outcomes. ICC estimates are summarised descriptively across trials by the median and range; no meta-analysis was performed. ICCs were calculated for the components of the composite outcomes for completeness, however, only summary information is presented for these outcomes as they are relatively rare and hence are unlikely to be chosen as the primary outcome for a future trial. Analyses were performed using SAS v9.4 (Cary, NC, USA) based on the %BOOT and %BOOTCI macros.[27]

RESULTS

Table 1 and Figure 1 summarise ICC estimates across trials for each of the 12 neonatal outcomes considered. ICCs were relatively high for the two composite measures of adverse neonatal outcome, with median (range) values of 0.68 (0.52-0.71) and 0.65 (0.54-0.77) across trials. For individual neonatal outcomes, median ICCs varied substantially from 0.13 for NEC to 0.79 for

NICU admission and birthweight. The vast majority of individual ICC estimates for each outcome and trial were above 0.5, indicating a moderate to strong correlation between adverse neonatal outcomes of twins. ICC estimates were generally fairly consistent across trials, despite considerable variation in outcome prevalence and differences in inclusion/exclusion criteria between trials. Chorionicity had no clear effect on ICC estimates, which were mostly similar for infants from monochorionic and dichorionic twin pregnancies (Appendix Tables S3-S8). Mixed effects models generally produced similar ICC estimates (Appendix Table S9).

Example Sample Size Calculation

To illustrate how the ICCs presented in this article can be used in sample size calculations for future RCTs in women with twin pregnancies, we present the following hypothetical example. Suppose a multicentre RCT is planned to assess the effect of a promising new drug for women with a twin pregnancy on adverse neonatal outcomes. Women with a monochorionic or dichorionic twin pregnancy diagnosed by ultrasound will be randomised between 16 and 20 weeks' gestation to receive the new drug or placebo in the ratio 1:1. The primary outcome for the trial is a composite neonatal outcome of perinatal death, RDS, BPD, IVH, NEC and sepsis. The outcome prevalence in the control group is expected to be 15% and the trial investigators believe the new drug will reduce the prevalence by at least 40%. Two steps are involved in calculating the sample size for RCTs in women with twin pregnancies. First, the sample size is calculated using standard methods assuming outcomes of infants from a twin pregnancy are independent. If the proposed trial were conducted under this assumption, a total of 986 infants (493 per group) would be required to detect a 40% reduction in the risk of adverse neonatal outcome from 15% to 9%, based on a continuity-

corrected chi-square test with two-sided $\alpha = 0.05$ and 80% power. Second, the sample size is multiplied by a quantity known as the design effect, which is given by 1+ICC for trials randomising and treating pregnant women and only including twin pregnancies.[28] The ICC estimates presented in this article can be used to calculate this design effect and hence the final sample size. The median ICC for the primary outcome of the proposed trial across previous similar trials is 0.68 (Table 1), which produces a design effect of 1.68 and increases the sample size for the proposed trial to a total of 1.68×986=1658 twin infants (after rounding up to the next even number), or 829 women with a twin pregnancy. Power calculations can be performed to examine the impact on power if the ICC is at the upper end of the range of likely values. For the proposed trial, the sample size of 1658 infants based on an ICC of 0.68 would provide 79% or 75% power if the ICC turned out to be 0.71 or 0.88 respectively, corresponding to the maximum values for the ICC estimate and the upper limit of the 95% confidence interval for the ICC estimate observed across similar trials (Appendix Table S3).

DISCUSSION

Main Findings

We present estimates of the correlation between outcomes of twins for a range of commonly reported neonatal outcomes using data from 12 RCTs randomising women with twin pregnancies. ICCs were generally above 0.5, indicating moderate to strong correlation between neonatal outcomes of twins, and were generally similar by chorionicity. ICCs were also fairly consistent across trials, despite differences in outcome prevalence and inclusion/exclusion criteria. However,

there was considerable variability in ICCs between outcomes and some ICCs were lower than may be expected for twins. Our example sample size calculation illustrates how these ICCs can be used in the design of RCTs in women with twin pregnancies and the large impact that twins can have on the sample size.

Strengths and Limitations

The key strength of this study is that, to our knowledge, it provides the first comprehensive report of ICCs for neonatal outcomes in twins. These ICCs will inform the design and analysis of future RCTs and systematic reviews evaluating interventions designed to improve neonatal outcomes in women with twin pregnancies. Another strength is the use of data from multiple RCTs to provide multiple estimates of the ICC for each outcome. This provides researchers with a range of likely ICC values for each neonatal outcome of interest.

A limitation of this study is that the ICCs were estimated from RCTs chosen for convenience, the vast majority of which investigated the effect of progestogens on neonatal outcomes, and may not be representative of all RCTs in women with twin pregnancies. Additional ICC estimates are needed from other RCTs and epidemiological studies involving twin pregnancies that focus on different clinical conditions and employ varying inclusion/exclusion criteria to obtain a more complete picture of the dependence between neonatal outcomes that occurs in twins. A further limitation is that we did not investigate the degree of outcome concordance within twin pairs that is beyond chance and this is an interesting area for further research.

Interpretation

External estimates of ICCs for neonatal outcomes in twins, such as those presented in this article, can be used by researchers in several settings. The most common use is likely to be in designing RCTs in women with twin pregnancies, where it is important to account for the dependence between neonatal outcomes of twins in sample size calculations to ensure the trial is adequately powered to answer the primary research question. This can be achieved by simply calculating the sample size using standard methods assuming outcomes of all infants are independent and then multiplying by a design effect of 1+ICC.[28] Our example sample size calculation illustrates this process using the median ICC across trials, although in practice it may be sensible to use the ICC estimate from the most similar trial in terms of inclusion/exclusion criteria. Alternatively, an ICC estimate may be obtained from a pilot study, although this requires resources that may not be available and is likely to yield a very imprecise ICC estimate. As our ICC estimates were generally above 0.5, this indicates that RCTs focusing on twins are likely to require at least 50% more infants than RCTs focusing on singletons, and that failure to account for twins in the sample size calculation could result in a trial with much lower than expected power. This does not necessarily mean that appropriately powered RCTs in twins will be more expensive than trials in singletons, however, as the costs associated with recruiting mothers and collecting mother level information are halved for twins. Many RCTs allow women with either a singleton or twin pregnancy to participate, and our ICC estimates can also be used to calculate the sample size for these trials by incorporating the twin pregnancy rate in the target population into the calculation of the design effect.[28]

Another likely use of external ICC estimates is in the analysis of RCTs including women with twin pregnancies. Previous studies have investigated the performance of different statistical methods for analysing neonatal outcomes in twins and recommended using an approach that takes the correlation between outcomes of twins into account, such as generalised estimating equations or mixed effects models.[10, 11, 29-32] If a trial is too small or includes too few women with a multiple pregnancy to provide a precise estimate of the ICC in the analysis, it may be preferable to use an external estimate. The Bayesian framework provides a formal method of incorporating external evidence into the analysis by specifying an informative prior for the ICC.[33] This has the advantage of utilising the uncertainty around the ICC estimate as well as the point value, and may be the most appropriate way to use the external information.

The final anticipated use of external ICC estimates is in systematic reviews and meta-analyses involving RCTs that include women with twin pregnancies. Adjustment of standard errors or sample size is common in meta-analyses of outcomes collected in cluster RCTs[34] but this approach is rarely applied to outcomes of infants from multiple pregnancies. By providing estimates of ICCs for neonatal outcomes in twins, we hope to encourage researchers to perform similar adjustments for meta-analyses including RCTs that recruited women with twin pregnancies. Such adjustments can appropriately increase the uncertainty around the treatment effect estimates and help guard against overly optimistic conclusions regarding the effectiveness of the intervention.

As expected, we found considerable variability in ICCs between neonatal outcomes. This variability may be due to differences in outcome prevalence, as well as the nature of the outcome. Median ICC estimates were as low as 0.13, which is substantially lower than we had anticipated

for neonatal outcomes of twins. As this median was based on only 2 trials with sufficient data to estimate the ICC for NEC, this finding should be interpreted with some caution. The next lowest median ICC estimates observed were 0.36 for IVH and 0.38 for sepsis, which are also somewhat lower than anticipated. We also expected ICCs to be higher for monochorionic compared to dichorionic twins due to the shared placenta, however chorionicity had no clear effect on ICC estimates. This could be due to the relatively small sample sizes available in these subgroups, as reflected in the wide confidence intervals for the ICCs, or unequal placental sharing in monochorionic twins. Alternatively, it may be due to the choice of neonatal outcomes studied, many of which are imprecise measures of the underlying clinical state. Further investigation of the impact of chorionicity on ICCs using data from larger epidemiological studies would be useful for informing the design and analysis of future RCTs specifically recruiting women with monochorionic or dichorionic twin pregnancies.

CONCLUSION

The correlation between neonatal outcomes of twins varies considerably between outcomes. It is generally moderate to high but may be lower than expected for some outcomes. This highlights the importance of obtaining an accurate estimate of the ICC for the relevant outcome and population to use in the design and analysis of RCTs that recruit women with twin pregnancies. Our ICC estimates will be useful to researchers requiring external information on these parameters for calculating the sample size, performing Bayesian analyses and adjusting meta-analyses to account for twins. Future RCTs including women with twin pregnancies should make use of these and other suitable ICC estimates during the trial design phase to ensure they are adequately

powered to answer the primary research question. Researchers are encouraged to report ICCs for neonatal outcomes in twins in their own trials to add to the growing body of published ICCs.

Acknowledgements

We are grateful to the trial investigators, research staff and participants who were involved with each of the trials contributing data to this study. In particular, we acknowledge the STOPPIT study team for the Norman trial; the MEDNAX Center for Research Education and Quality and use of its data from the Combs trial; the assistance of the NICHD, the MFMU Network and the STTARS Protocol Subcommittee in making the database available for the Rouse trial; and the following trial investigators who granted permission to use data from their trials: Jane Norman, Stephen Wood, Mona Aboulghar, Kimberly Maurel, Sophie Liem and Elcin Cetingoz. The contents of this report represent the views of the authors and do not represent the views of the Eunice Kennedy Shriver National Institute of Child Health and Human Development Maternal-Fetal Medicine Units Network or the National Institutes of Health.

Disclosure of interests

None of the authors has any financial, personal, political, academic, or other relationships that could lead to a conflict of interest relevant to this article.

Contribution to authorship

LNY, ES, JZ, PM, BWJM and SG were involved in the concept and the design of the study. ACL, AHN, LR, VS, ET and CV contributed data to the study and ES provided the datasets and related support. LNY performed the analyses and drafted the initial manuscript, with significant contributions by ES, JZ, PM, BWJM and SG. All authors critically reviewed the manuscript and approved the final version for submission.

Details of ethical approval

All trials had institutional review board approval and informed consent from all participants. Ethical approval was obtained for this study from the Women's & Children's Health Network Human Research Ethics Committee (HREC/14/WCHN/165, December 2014).

Funding

Lisa Yelland was supported by an Australian National Health and Medical Research Council Early Career Fellowship (ID 1052388). This research was supported by a Women's and Children's Hospital Foundation Research Project Grant. The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Table S1. Characteristics of Trials Used to Estimate Intraclass Correlation Coefficients

Table S2. Sample Size by Trial and Chorionicity

Table S3. Intraclass Correlation Coefficients for Composite Adverse Neonatal Outcome 1 by Trial and Chorionicity

Table S4. Intraclass Correlation Coefficients for Composite Adverse Neonatal Outcome 2 by Trial and Chorionicity

Table S5. Intraclass Correlation Coefficients for Admission to Neonatal Intensive Care Unit by Trial and Chorionicity

Table S6. Intraclass Correlation Coefficients for Birthweight by Trial and Chorionicity

Table S7. Intraclass Correlation Coefficients for Birthweight <2500g by Trial and Chorionicity

Table S8. Intraclass Correlation Coefficients for Birthweight <1500g by Trial and Chorionicity

Table S9. Summary of Intraclass Correlation Coefficient Estimates for Neonatal Outcomes from Linear Mixed Effects Models Across Trials

References

1. Ananth CV, Chauhan SP. Epidemiology of twinning in developed countries. Seminars in Perinatology. 2012 Jun;36(3):156-61.
2. Blondel B, Kogan MD, Alexander GR, Dattani N, Kramer MS, Macfarlane A, et al. The impact of the increasing number of multiple births on the rates of preterm birth and low birthweight: an international study. American Journal of Public Health. 2002 Aug;92(8):1323-30.
3. Shinwell ES, Haklai T, Eventov-Friedman S. Outcomes of multiplets. Neonatology. 2009;95(1):6-14.
4. Rode L, Klein K, Nicolaides KH, Krampl-Bettelheim E, Tabor A, Group P. Prevention of preterm delivery in twin gestations (PREDICT): a multicenter, randomized, placebo-controlled trial on the effect of vaginal micronized progesterone. Ultrasound in Obstetrics and Gynecology. 2011 Sep;38(3):272-80.
5. Norman JE, Mackenzie F, Owen P, Mactier H, Hanretty K, Cooper S, et al. Progesterone for the prevention of preterm birth in twin pregnancy (STOPPIT): a randomised, double-blind, placebo-controlled study and meta-analysis. Lancet. 2009 Jun 13;373(9680):2034-40.
6. Serra V, Perales A, Meseguer J, Parrilla JJ, Lara C, Bellver J, et al. Increased doses of vaginal progesterone for the prevention of preterm birth in twin pregnancies: a randomised controlled double-blind multicentre trial. BJOG: an International Journal of Obstetrics and Gynaecology. 2013 Jan;120(1):50-7.
7. Awwad J, Usta IM, Ghazeeri G, Yacoub N, Succar J, Hayek S, et al. A randomised controlled double-blind clinical trial of 17-hydroxyprogesterone caproate for the

prevention of preterm birth in twin gestation (PROGESTWIN): evidence for reduced neonatal morbidity. BJOG: an International Journal of Obstetrics and Gynaecology. 2015 Jan;122(1):71-9.

8. Combs CA, Garite T, Maurel K, Das A, Porto M, Obstetrix Collaborative Research N. 17-hydroxyprogesterone caproate for twin pregnancy: a double-blind, randomized clinical trial. American Journal of Obstetrics and Gynecology. 2011 Mar;204(3):221 e1-8.

9. Senat MV, Porcher R, Winer N, Vayssiere C, Deruelle P, Capelle M, et al. Prevention of preterm delivery by 17 alpha-hydroxyprogesterone caproate in asymptomatic twin pregnancies with a short cervix: a randomized controlled trial. American Journal of Obstetrics and Gynecology. 2013 Mar;208(3):194 e1-8.

10. Gates S, Brocklehurst P. How should randomised trials including multiple pregnancies be analysed? BJOG: an International Journal of Obstetrics and Gynaecology. 2004 Mar;111(3):213-9.

11. Marston L, Peacock JL, Yu KM, Brocklehurst P, Calvert SA, Greenough A, et al. Comparing methods of analysing datasets with small clusters: case studies using four paediatric datasets. Paediatric and Perinatal Epidemiology. 2009 Jul;23(4):380-92.

12. Killip S, Mahfoud Z, Pearce K. What is an intracluster correlation coefficient? Crucial concepts for primary care researchers. Annals of Family Medicine. 2004 May-Jun;2(3):204-8.

13. Campbell MK, Piaggio G, Elbourne DR, Altman DG, Group C. Consort 2010 statement: extension to cluster randomised trials. British Medical Journal. 2012 Sep 4;345:e5661.

14. Yelland LN, Sullivan TR, Makrides M. Accounting for multiple births in randomised trials: a systematic review. Archives of Disease in Childhood-Fetal and Neonatal Edition. 2015 Mar;100(2):F116-20.

15. Liem S, Schuit E, Hegeman M, Bais J, de Boer K, Bloemenkamp K, et al. Cervical pessaries for prevention of preterm birth in women with a multiple pregnancy (ProTWIN): a multicentre, open-label randomised controlled trial. Lancet. 2013 Oct 19;382(9901):1341-9.

16. Schuit E, Stock S, Rode L, Rouse DJ, Lim AC, Norman JE, et al. Effectiveness of progestogens to improve perinatal outcome in twin pregnancies: an individual participant data meta-analysis. BJOG: an International Journal of Obstetrics and Gynaecology. 2015 Jan;122(1):27-37.

17. Rouse DJ, Caritis SN, Peaceman AM, Sciscione A, Thom EA, Spong CY, et al. A trial of 17 alpha-hydroxyprogesterone caproate to prevent prematurity in twins. New Engl J Med. 2007 Aug 02;357(5):454-61.

18. Lim AC, Schuit E, Bloemenkamp K, Bernardus RE, Duvekot JJ, Erwich JJ, et al. 17alpha-hydroxyprogesterone caproate for the prevention of adverse neonatal outcome in multiple pregnancies: a randomized controlled trial. Obstetrics and Gynecology. 2011 Sep;118(3):513-20.

19. Aboulghar MM, Aboulghar MA, Amin YM, Al-Inany HG, Mansour RT, Serour GI. The use of vaginal natural progesterone for prevention of preterm birth in IVF/ICSI pregnancies. Reproductive Biomedicine Online. 2012;25:133-8.

20. Wood S, Ross S, Tang S, Miller L, Sauve R, Brant R. Vaginal progesterone to prevent preterm birth in multiple pregnancy: a randomized controlled trial. Journal of Perinatal Medicine 2012 Nov;40(6):593-9.

21. Cetingoz E, Cam C, Sakalli M, Karateke A, Celik C, Sancak A. Progesterone effects on preterm birth in high-risk pregnancies: a randomized placebo-controlled trial. Archives of Gynecology and Obstetrics. 2011 Mar;283(3):423-9.

22. Giraudeau B. Model mis-specification and overestimation of the intraclass correlation coefficient in cluster randomized trials. Statistics in Medicine. 2006 Mar 30;25(6):957-64.

23. Hibbs AM, Black D, Palermo L, Cnaan A, Luan XQ, Truog WE, et al. Accounting for multiple births in neonatal and perinatal trials: systematic review and case study. Journal of Pediatrics. 2010 Feb;156(2):202-8.

24. Carpenter J, Bithell J. Bootstrap confidence intervals: when, which, what? A practical guide for medical statisticians. Statistics in Medicine. 2000 May 15;19(9):1141-64.

25. Lu B, Preisser JS, Qaqish BF, Suchindran C, Bangdiwala S, Wolfson M. A comparison of two bias-corrected covariance estimators for generalized estimating equations. Biometrics. 2007 Sep;63(3):935-41.

26. Mancl LA, DeRouen TA. A covariance estimator for GEE with improved small-sample properties. Biometrics. 2001 Mar;57(1):126-34.

27. SAS Institute Inc. jackboot.sas. [cited 03/07/2008]; Available from: http://support.sas.com/kb/24/982.html

28. Yelland LN, Sullivan TR, Price DJ, Lee KJ. Sample size calculations for randomised trials including both independent and paired data. Statistics in Medicine. 2017 Apr 15;36(8):1227-39.

29. Ananth CV, Platt RW, Savitz DA. Regression models for clustered binary responses: implications of ignoring the intracluster correlation in an analysis of perinatal mortality in twin gestations. Annals of Epidemiology. 2005 Apr;15(4):293-301.

30. Sauzet O, Peacock JL. Binomial outcomes in dataset with some clusters of size two: can the dependence of twins be accounted for? A simulation study comparing the reliability of statistical methods based on a dataset of preterm infants. Bmc Med Res Methodol. 2017 Jul 20;17.

31. Sauzet O, Wright KC, Marston L, Brocklehurst P, Peacock JL. Modelling the hierarchical structure in datasets with very small clusters: a simulation study to explore the effect of the proportion of clusters when the outcome is continuous. Statistics in Medicine. 2013 Apr 15;32(8):1429-38.

32. Yelland LN, Salter AB, Ryan P, Makrides M. Analysis of binary outcomes from randomised trials including multiple births: when should clustering be taken into account? Paediatric and Perinatal Epidemiology. 2011;25:283-97.

33. Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB. Bayesian data analysis. Third edition. ed. Boca Raton: CRC Press; 2014.

34. Higgins JPT, Green S, (editors). Cochrane Handbook for Systematic Reviews of Interventions. www.handbook.cochrane.org: The Cochrane Collaboration; 2011.

Table 1. Summary of Intraclass Correlation Coefficient Estimates for Neonatal Outcomes Across Trials

Outcome	Median (Range) ICC	Trials

Composite Adverse Neonatal Outcome 1[a]	0.68 (0.52-0.71)	5-9, 15, 17, 18, 21
Composite Adverse Neonatal Outcome 2[b]	0.65 (0.54-0.77)	4-9, 15, 17, 18, 20, 21
Perinatal Death	0.66 (0.17-0.80)	4, 5, 7, 15, 17, 18, 21
Respiratory Distress Syndrome	0.65 (0.50-0.74)	4-9, 15, 17, 18, 20, 21
Bronchopulmonary Dysplasia	0.51 (0.37-0.72)	5, 17, 18
Intraventricular Haemorrhage	0.36 (0.13-0.45)	4, 5, 17
Necrotising Enterocolitis	0.13 (0.12-0.14)	15, 18
Sepsis	0.38 (0.35-0.47)	4, 5, 7, 15, 17, 18
Admission to Neonatal Intensive Care Unit	0.79 (0.56-0.86)	4-9, 15, 17, 18, 21
Birthweight	0.79 (0.62-0.85)	4-9, 15, 17-21
Birthweight <2500g	0.50 (0.37-0.71)	4-9, 15, 17-21
Birthweight <1500g	0.71 (0.36-0.91)	4-9, 15, 17, 18, 20, 21

[a] Includes perinatal death, respiratory distress syndrome, bronchopulmonary dysplasia, intraventricular haemorrhage, necrotising enterocolitis and sepsis

[b] Includes perinatal death, respiratory distress syndrome, intraventricular haemorrhage and necrotising enterocolitis

Figure 1. Boxplots of intraclass correlation coefficient estimates across trials by outcome. Abbreviations: COMP, composite adverse neonatal outcome; Death, perinatal death; RDS, respiratory distress syndrome; BPD, bronchopulmonary dysplasia; IVH, intraventricular

haemorrhage; NEC, necrotizing enterocolitis; NICU, neonatal intensive care unit; BW, birthweight; ICC, intraclass correlation coefficient.

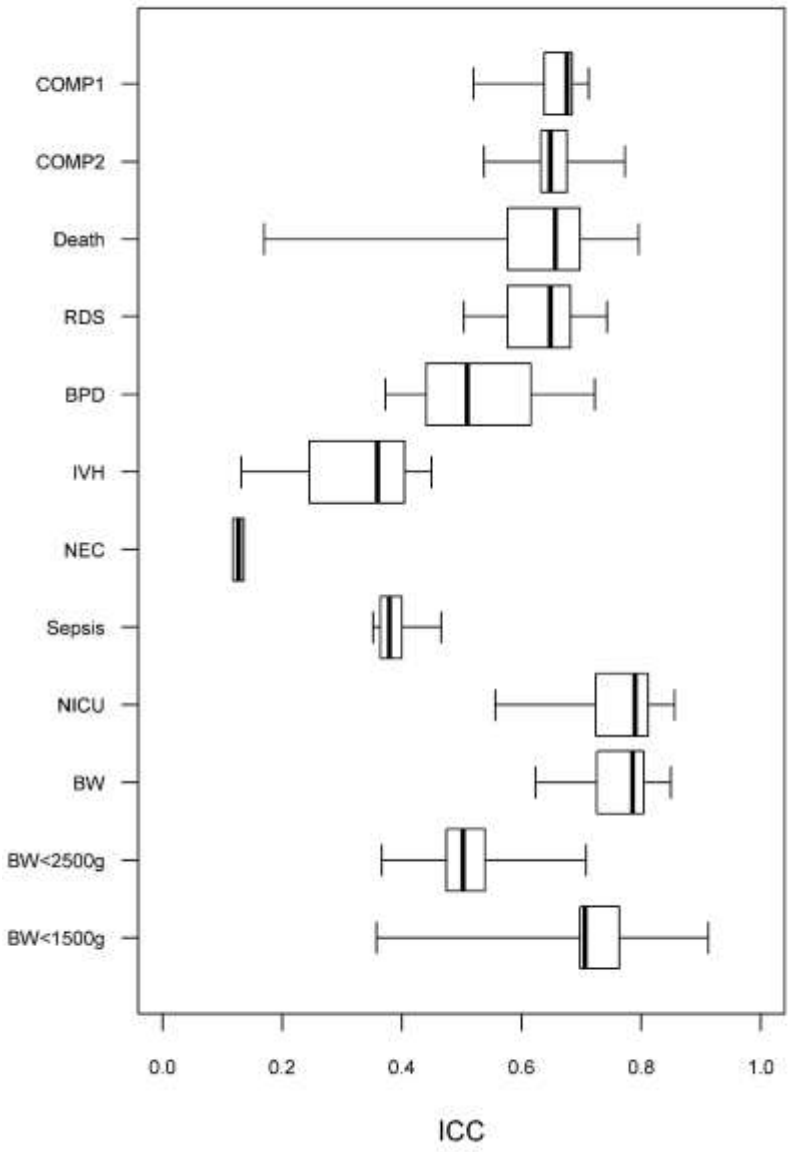

Table S1. Characteristics of Trials Used to Estimate Intraclass Correlation Coefficients

Trial	Study Design	Inclusion Criteria	Exclusion Criteria	Treatment Groups
Cervical Pessary ProTWIN Trial (Liem)[15]	Multicentre, open-label RCT	Women with a multiple pregnancy 12-20 weeks' gestation	Known serious congenital defects, fetal death, twin-to-twin transfusion syndrome, known placenta previa	Cervical pessary inserted 16-20 weeks' gestation and removed in the 36th week of gestation vs no cervical pessary
Progestogen Individual Patient Data Meta-Analysis				
- Rode[4]	Multicentre, double-blind, placebo-controlled RCT	Women with a live, diamniotic twin pregnancy and chorionicity assessed by ultrasound <16 weeks' gestation	Age <18 years, known allergy to progesterone or peanuts, history of hormone-associated thromboembolic disorders, rupture of membranes, treatment for signs of twin-to-twin transfusion syndrome, intentional fetal reduction, known major structural or chromosomal fetal abnormality, known or suspected malignancy in genitals or breasts, known liver disease, higher-order multiple pregnancies, women who did not speak and understand Danish or German, as appropriate	Vaginal progesterone pessaries (200mg) vs vaginal placebo pessaries self-administered daily from 20+0-23+6 weeks' gestation until 33+6 weeks' or occurrence of either rupture of membranes or delivery
- Rouse[17]	Multicentre, double-blind, placebo-controlled RCT	Women carrying twins 16+0-20+3 weeks' gestation	Serious fetal anomalies, spontaneous death of a fetus >12 weeks, presumed monoamniotic placenta, suspected twin-to-twin transfusion syndrome, marked ultrasonographic growth discordance, planned nonstudy progesterone therapy >16 weeks, in-place or planned cerclage, major uterine anomaly, treatment with	Weekly intramuscular injections of 17Pc (250mg) vs placebo starting at 16+0-20+6 weeks' gestation and continuing until the end of the 34th week of gestation or delivery

			10,000 or more units of unfractionated heparin per day, treatment with low-molecular-weight heparin at any dose, major chronic medical diseases, twin gestations that were the result of intentional fetal reduction	
- Lim[18]	Multicentre, double-blind, placebo-controlled RCT	Women with a multiple pregnancy 15-19 weeks' gestation and chorionicity determined by ultrasonography	Women with a previous spontaneous preterm birth <34 weeks, serious congenital defects or death of one or more fetuses, early signs of twin-to-twin transfusion syndrome, primary cerclage	Weekly intramuscular injections of 17Pc (250mg) vs placebo from 16-20 weeks' gestation until 36 weeks' or delivery
- Norman[5]	Multicentre, double-blind, placebo-controlled RCT	Women with a twin pregnancy, with gestation and chorionicity established by scan <20 weeks' gestation, and attending the antenatal clinic during the recruitment period	Pregnancy complicated by a recognised structural or chromosomal fetal abnormality at the time of recruitment, contraindications to progesterone, planned cervical suture, planned elective delivery <34 weeks, planned intervention for twin-to-twin transfusion <22 weeks, higher order multiple pregnancy	Daily progesterone gel (90mg) vs placebo self-administered vaginally for 10 weeks from 24+0 weeks' gestation
- Serra[6]	Multicentre, double-blind, placebo-controlled RCT	Maternal age ≥18 years, dichorionic diamniotic twin pregnancy diagnosed by ultrasound and written informed consent	Singleton pregnancies, monochorionic twin pregnancies, triplets or higher order multiple pregnancies, elective cervical cerclage <14 weeks, history of hepatic problems or gestational cholestasis, abnormal liver enzymes, abnormal kidney function, local allergy to micronised natural progesterone, allergy to peanuts, recurrent vaginal bleeding, recurrent vaginal infections, fetal anomalies diagnosed by ultrasound, alcohol or illicit	Two vaginal progesterone pessaries (400mg or 200mg) vs placebo self-inserted daily at bedtime from 20 weeks' gestation until 34 weeks' or delivery

			drug consumption, smoking ≥10 cigarettes/day	
- Nassar[7]	Single centre, double-blind, placebo-controlled RCT	Twin pregnancy diagnosed by ultrasound and maternal age ≥18 years, recruited at 12-20 weeks' gestation	Ultrasonographically diagnosed fetal anomalies, elective cervical cerclage <14 weeks, hypertension, diabetes mellitus, asthma, history of deep vein thrombosis, history of hepatic disease or abnormal liver enzymes, pre-existing renal disease or abnormal kidney function, seizure disorders	Weekly intramuscular injections of 17Pc (250mg) vs placebo from16-20 weeks' gestation until 36 weeks'
- Combs[8]	Multicentre, double-blind, placebo-controlled RCT	Women with a dichorionic-diamniotic twin pregnancy at 15-23 weeks' gestation with a detailed ultrasound examination showing no major fetal anomalies	Age <18 years, taken any progestins >15 weeks, symptomatic uterine contractions, rupture of fetal membranes, contraindication to prolonging the pregnancy, pre-existing condition that might be worsened by progesterone, pre-existing medical condition carrying a high risk of preterm delivery	Weekly intramuscular injections of 17Pc (250mg) vs placebo from 16-24 weeks' gestation until 34 weeks' or delivery
- Senat[9]	Multicentre, open-label RCT	Women >18 years, carrying twins, asymptomatic, cervical length ≤25mm measured in the sagittal plane by routine transvaginal ultrasound according to the standard technique, who agreed to regular follow-up and provided written informed consent	Cervical dilatation >3cm, premature rupture of the membranes, placenta previa, monochorial monoamniotic pregnancy, signs of twin-to-twin transfusion syndrome, severe intrauterine growth restriction, known major structural or chromosomal fetal abnormality, death of 1 fetus, any maternal or fetal disease requiring preterm delivery, progesterone therapy before inclusion, ongoing anticonvulsant treatment, participation in any other treatment trial, twin gestations resulting from intentional fetal reduction	Twice weekly intramuscular injections of 17Pc (500mg) from 24+0-31+6 weeks' gestation until 36 weeks' or preterm delivery vs no treatment
- Aboulghar[19]	Single centre, placebo-	Healthy pregnant women who conceived	Previous pregnancy, serious fetal anomalies for which termination may be	Vaginal progesterone suppositories (200mg) vs

	controlled RCT	after IVF/ICSI between 18-24 weeks' gestation, with a first pregnancy, singleton or dichorionic twins, normal uterine and cervical anatomy, and normal fetal anatomy	considered, intrauterine growth restriction, mono-chorionic and mono-amniotic twins, uterine anomalies, triplet pregnancies, cervical cerclage	placebo twice daily from randomisation until 37 weeks' gestation or onset of preterm birth
- Wood[20]	Multicentre, double-blind, placebo-controlled RCT	Pregnant women with two or more live fetuses confirmed at 16-18 week ultrasound, 16+0-20+6 weeks' gestation	Placenta previa, pre-existing hypertension, known major fetal anomaly detected on ultrasound, monoamniotic monozygotic multiple pregnancies, maternal seizure disorder, active or history of thromboembolic disease, maternal liver disease, known or suspected breast malignancy or pathology, known or suspected progesterone-dependent neoplasia, plans to move to another city during pregnancy, previous participation in this trial or other perinatal clinical trials during this pregnancy, known sensitivity to progesterone	Daily progesterone gel (90mg) vs placebo self-administered vaginally from randomisation until 35+6 weeks' gestation
- Cetingoz[21]	Single centre, double-blind, placebo-controlled RCT	Women with a twin pregnancy, prior spontaneous preterm birth or uterine malformation	Abortions and deliveries 20-24 weeks, prophylactic cervical cerclage	Vaginal progesterone suppositories (100mg) vs placebo nightly from 24 weeks' gestation until 34 weeks'

Table S2. Sample Size by Trial and Chorionicity

Trial	Number of Women With a Twin Pregnancy[a]	Number (%) of Women with Monochorionic Pregnancy	Number (%) of Women with Dichorionic Pregnancy	Number (%) of Women with Unknown Chorionicity
Liem	795	181 (22.8)	609 (76.6)	5 (0.6)
Rode	677	100 (14.8)	577 (85.2)	0 (0.0)
Rouse	661	103 (15.6)	551 (83.4)	7 (1.1)
Lim	650	112 (17.2)	538 (82.8)	0 (0.0)
Norman	500	92 (18.4)	408 (81.6)	0 (0.0)
Serra	290	0 (0.0)	290 (100.0)	0 (0.0)
Nassar	286	41 (14.3)	222 (77.6)	23 (8.0)
Combs	240	0 (0.0)	240 (100.0)	0 (0.0)
Senat	165	0 (0.0)	0 (0.0)	165 (100.0)
Aboulghar	92	0 (0.0)	92 (100.0)	0 (0.0)
Wood	81	0 (0.0)	0 (0.0)	81 (100.0)
Cetingoz	67	9 (13.4)	26 (38.8)	32 (47.8)

[a] Some trials included women with single or higher order multiple pregnancies but only women with twin pregnancies were included in this study

Table S3. Intraclass Correlation Coefficients for Composite Adverse Neonatal Outcome 1[d] by Trial and Chorionicity

Trial	Prevalence (%)	ICC (95% CI) - All Twins	ICC (95% CI) - Monochorionic Twins	ICC (95% CI) - Dichorionic Twins
Liem	9.94	0.68 (0.59, 0.76)	0.62 (0.43, 0.78)	0.73 (0.62, 0.82)
Rouse	17.70	0.70 (0.62, 0.77)	0.86 (0.70, 0.96)	0.65 (0.56, 0.73)
Lim	15.25	0.68 (0.59, 0.75)	0.76 (0.56, 0.90)	0.66 (0.57, 0.75)
Norman	12.09	0.54 (0.43, 0.67)	0.50 (0.23, 0.77)	0.56 (0.41, 0.70)
Serra	14.66	0.52 (0.38, 0.66)	[a]	0.52 (0.38, 0.66)
Nassar	22.28	0.68 (0.56, 0.78)	0.86 (0.43, 1.00)	0.68 (0.55, 0.79)
Combs	14.04	0.71 (0.56, 0.84)	[a]	0.71 (0.56, 0.84)
Senat	29.93	0.64 (0.49, 0.77)	[b]	[b]
Cetingoz	17.16	0.65 (0.32, 0.88)	[c]	[c]

[a] Monochorionic twins excluded from trial

[b] Chorionicity unknown

[c] Insufficient data to estimate ICC

[d] Includes perinatal death, respiratory distress syndrome, bronchopulmonary dysplasia, intraventricular haemorrhage, necrotising enterocolitis and sepsis

Table S4. Intraclass Correlation Coefficients for Composite Adverse Neonatal Outcome 2[d] by Trial and Chorionicity

Trial	Prevalence (%)	ICC (95% CI) - All Twins	ICC (95% CI) - Monochorionic Twins	ICC (95% CI) - Dichorionic Twins
Liem	8.23	0.65 (0.54, 0.74)	0.63 (0.42, 0.82)	0.66 (0.54, 0.77)
Rode	11.87	0.77 (0.69, 0.84)	0.82 (0.57, 0.96)	0.76 (0.66, 0.84)
Rouse	17.08	0.68 (0.60, 0.75)	0.85 (0.69, 0.96)	0.62 (0.52, 0.71)
Lim	14.10	0.67 (0.59, 0.76)	0.80 (0.56, 0.93)	0.65 (0.55, 0.75)
Norman	10.85	0.56 (0.44, 0.68)	0.62 (0.31, 0.86)	0.54 (0.40, 0.68)
Serra	14.31	0.54 (0.40, 0.68)	[a]	0.54 (0.40, 0.68)
Nassar	20.53	0.67 (0.54, 0.77)	0.94 (0.36, 1.00)	0.64 (0.50, 0.76)
Combs	14.04	0.71 (0.56, 0.84)	[a]	0.71 (0.56, 0.84)
Senat	29.22	0.62 (0.46, 0.74)	[b]	[b]
Wood	20.37	0.65 (0.40, 0.87)	[b]	[b]
Cetingoz	17.16	0.65 (0.32, 0.88)	[c]	[c]

[a] Monochorionic twins excluded from trial

[b] Chorionicity unknown

[c] Insufficient data to estimate ICC

[d] Includes perinatal death, respiratory distress syndrome, intraventricular haemorrhage and necrotising enterocolitis

Table S5. Intraclass Correlation Coefficients for Admission to Neonatal Intensive Care Unit by Trial and Chorionicity

Trial	Prevalence (%)	ICC (95% CI) - All Twins	ICC (95% CI) - Monochorionic Twins	ICC (95% CI) - Dichorionic Twins
Liem	13.05	0.72 (0.64, 0.79)	0.67 (0.51, 0.80)	0.75 (0.66, 0.83)
Rode	48.82	0.86 (0.81, 0.89)	0.95 (0.85, 1.00)	0.84 (0.79, 0.88)
Rouse	48.58	0.81 (0.76, 0.85)	0.88 (0.74, 0.96)	0.80 (0.74, 0.85)
Lim	18.31	0.79 (0.72, 0.85)	0.84 (0.69, 0.94)	0.77 (0.69, 0.84)
Norman	39.40	0.79 (0.73, 0.84)	0.87 (0.74, 0.96)	0.77 (0.70, 0.83)
Serra	11.90	0.56 (0.40, 0.71)	[a]	0.56 (0.40, 0.71)
Nassar	36.89	0.80 (0.72, 0.87)	[c]	0.76 (0.66, 0.84)
Combs	38.56	0.79 (0.70, 0.86)	[a]	0.79 (0.70, 0.86)
Senat	41.21	0.85 (0.75, 0.93)	[b]	[b]
Cetingoz	29.10	0.68 (0.42, 0.86)	[c]	[c]

[a] Monochorionic twins excluded from trial

[b] Chorionicity unknown

[c] Insufficient data to estimate ICC

Table S6. Intraclass Correlation Coefficients for Birthweight by Trial and Chorionicity

Trial	Mean (SD)	ICC (95% CI) - All Twins	ICC (95% CI) - Monochorionic Twins	ICC (95% CI) - Dichorionic Twins
Liem	2344 (637)	0.81 (0.77, 0.83)	0.83 (0.76, 0.88)	0.80 (0.75, 0.83)
Rode	2434 (584)	0.80 (0.76, 0.83)	0.85 (0.77, 0.91)	0.79 (0.74, 0.83)
Rouse	2259 (617)	0.85 (0.82, 0.87)	0.88 (0.83, 0.93)	0.84 (0.80, 0.86)
Lim	2362 (683)	0.80 (0.75, 0.84)	0.78 (0.63, 0.88)	0.81 (0.74, 0.85)
Norman	2325 (619)	0.79 (0.74, 0.83)	0.85 (0.76, 0.91)	0.78 (0.72, 0.82)
Serra	2350 (508)	0.70 (0.62, 0.77)	[a]	0.70 (0.62, 0.77)
Nassar	2241 (569)	0.78 (0.71, 0.83)	0.79 (0.62, 0.91)	0.77 (0.69, 0.83)
Combs	2371 (534)	0.70 (0.60, 0.77)	[a]	0.70 (0.60, 0.77)
Senat	2145 (534)	0.83 (0.77, 0.88)	[b]	[b]
Aboulghar	2345 (505)	0.62 (0.46, 0.77)	[a]	0.62 (0.46, 0.77)
Wood	2291 (559)	0.75 (0.62, 0.86)	[b]	[b]
Cetingoz	2288 (562)	0.78 (0.59, 0.89)	[c]	[c]

[a] Monochorionic twins excluded from trial

[b] Chorionicity unknown

[c] Insufficient data to estimate ICC

Table S7. Intraclass Correlation Coefficients for Birthweight <2500g by Trial and Chorionicity

Trial	Prevalence (%)	ICC (95% CI) - All Twins	ICC (95% CI) - Monochorionic Twins	ICC (95% CI) - Dichorionic Twins
Liem	54.87	0.50 (0.44, 0.56)	0.47 (0.32, 0.60)	0.50 (0.43, 0.57)
Rode	49.85	0.50 (0.43, 0.57)	0.61 (0.44, 0.78)	0.48 (0.41, 0.55)
Rouse	61.95	0.61 (0.54, 0.67)	0.58 (0.35, 0.77)	0.60 (0.53, 0.67)
Lim	51.85	0.52 (0.45, 0.59)	0.48 (0.31, 0.65)	0.53 (0.45, 0.60)
Norman	56.69	0.48 (0.40, 0.56)	0.41 (0.21, 0.61)	0.49 (0.41, 0.58)
Serra	57.96	0.47 (0.36, 0.57)	[a]	0.47 (0.36, 0.57)
Nassar	64.57	0.53 (0.43, 0.64)	0.45 (0.15, 0.72)	0.53 (0.41, 0.66)
Combs	55.49	0.50 (0.39, 0.62)	[a]	0.50 (0.39, 0.62)
Senat	74.68	0.55 (0.39, 0.70)	[b]	[b]
Aboulghar	51.95	0.41 (0.20, 0.61)	[a]	0.41 (0.20, 0.61)
Wood	56.88	0.37 (0.16, 0.58)	[b]	[b]
Cetingoz	56.72	0.71 (0.51, 0.88)	[c]	[c]

[a] Monochorionic twins excluded from trial

[b] Chorionicity unknown

[c] Insufficient data to estimate ICC

Table S8. Intraclass Correlation Coefficients for Birthweight <1500g by Trial and Chorionicity

Trial	Prevalence (%)	ICC (95% CI) - All Twins	ICC (95% CI) - Monochorionic Twins	ICC (95% CI) - Dichorionic Twins
Liem	9.42	0.75 (0.66, 0.83)	0.71 (0.52, 0.89)	0.76 (0.65, 0.84)
Rode	6.72	0.78 (0.68, 0.87)	0.44 (-0.03, 0.80)	0.80 (0.68, 0.89)
Rouse	11.00	0.77 (0.68, 0.85)	0.72 (0.50, 0.88)	0.79 (0.69, 0.87)
Lim	10.96	0.75 (0.66, 0.82)	0.90 (0.74, 1.00)	0.72 (0.62, 0.81)
Norman	9.73	0.70 (0.58, 0.80)	0.73 (0.36, 0.93)	0.71 (0.57, 0.83)
Serra	6.06	0.48 (0.29, 0.72)	[a]	0.48 (0.29, 0.72)
Nassar	9.89	0.70 (0.51, 0.84)	[c]	0.73 (0.55, 0.87)
Combs	7.59	0.71 (0.45, 0.87)	[a]	0.71 (0.45, 0.87)
Senat	14.29	0.70 (0.38, 0.88)	[b]	[b]
Wood	10.00	0.91 (0.63, 1.00)	[b]	[b]
Cetingoz	8.21	0.36 (-0.05, 0.92)	[c]	[c]

[a] Monochorionic twins excluded from trial

[b] Chorionicity unknown

[c] Insufficient data to estimate ICC

Table S9. Summary of Intraclass Correlation Coefficient Estimates for Neonatal Outcomes from Linear Mixed Effects Models Across Trials

Outcome	Median (Range) ICC	Trials
Composite Adverse Neonatal Outcome 1[a]	0.68 (0.54-0.71)	5-9, 15, 17, 18, 21
Composite Adverse Neonatal Outcome 2[b]	0.66 (0.56-0.77)	4-9, 15, 17, 18, 20, 21
Perinatal Death	0.67 (0.16-0.79)	4, 5, 7, 15, 17, 18, 21
Respiratory Distress Syndrome	0.65 (0.50-0.74)	4-9, 15, 17, 18, 20, 21
Bronchopulmonary Dysplasia	0.51 (0.36-0.72)	5, 17, 18
Intraventricular Haemorrhage	0.37 (0.15-0.46)	4, 5, 17
Necrotising Enterocolitis	0.14 (0.14-0.15)	15, 18
Sepsis	0.40 (0.35-0.51)	4, 5, 7, 15, 17, 18
Admission to Neonatal Intensive Care Unit	0.79 (0.56-0.86)	4-9, 15, 17, 18, 21
Birthweight	0.78 (0.62-0.85)	4-9, 15, 17-21
Birthweight <2500g	0.50 (0.37-0.70)	4-9, 15, 17-21
Birthweight <1500g	0.72 (0.31-0.86)	4-9, 15, 17, 18, 20, 21

[a] Includes perinatal death, respiratory distress syndrome, bronchopulmonary dysplasia, intraventricular haemorrhage, necrotising enterocolitis and sepsis

[b] Includes perinatal death, respiratory distress syndrome, intraventricular haemorrhage and necrotising enterocolitis