



Published in final edited form as:

BJOG. 2018 October ; 125(11): 1406–1413. doi:10.1111/1471-0528.15292.

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

LN Yelland<sup>a,b</sup>, E Schuit<sup>c</sup>, J Zamora<sup>d,e</sup>, PF Middleton<sup>a,f</sup>, AC Lim<sup>g</sup>, AH Nassar<sup>h</sup>, L Rode<sup>i,j</sup>, V Serra<sup>k</sup>, EA Thom<sup>l</sup>, C Vayssi re<sup>m,n</sup>, BWJ Mol<sup>o</sup>, S Gates<sup>p</sup>

<sup>a</sup>South Australian Health and Medical Research Institute, Adelaide, SA, Australia <sup>b</sup>School of Public Health, The University of Adelaide, Adelaide, SA, Australia <sup>c</sup>Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, the Netherlands <sup>d</sup>Women’s Health Research Unit, Barts and the London School of Medicine and Dentistry, Queen Mary University of London, London, UK <sup>e</sup>CIBER Epidemiology and Public Health and IRYCIS, Madrid, Spain <sup>f</sup>Robinson Research Institute, School of Medicine, The University of Adelaide, Adelaide, SA, Australia <sup>g</sup>Department of Obstetrics & Gynecology and GROW-School for Oncology and Developmental Biology, Maastricht University Medical Centre, Maastricht, the Netherlands <sup>h</sup>Department of Obstetrics and Gynecology, American University of Beirut Medical Center, Beirut, Lebanon <sup>i</sup>Department of Obstetrics, Rigshospitalet, Copenhagen University Hospital, Copenhagen, Denmark <sup>j</sup>Department of Clinical Biochemistry, Herlev and Gentofte Hospital, Copenhagen University Hospital, Copenhagen, Denmark <sup>k</sup>Maternal-Fetal Medicine Unit, Valencia Infertility Institute, University of Valencia, Valencia, Spain <sup>l</sup>George Washington University Biostatistics Center, Washington, DC, USA <sup>m</sup>Department of Obstetrics and Gynecology, Paule de Viguier Hospital, CHU Toulouse, Toulouse, France <sup>n</sup>UMR 1027 INSERM, University of Paul Sabatier Toulouse III, Toulouse, France <sup>o</sup>Department of Obstetrics and Gynaecology, Monash Medical Centre, Monash Health and Monash University, Clayton, Vic., Australia <sup>p</sup>Warwick Clinical Trials Unit, University of Warwick, Coventry, UK

### Abstract

*Correspondence:* Dr LN Yelland, School of Public Health, Mail Drop DX650 511, The University of Adelaide, SA 5005, Australia. lisa.yelland@adelaide.edu.au.

#### Contribution to authorship

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

#### Disclosure of interests

Full disclosure of interests available to view online as supporting information.

#### Details of ethical approval

All trials received 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).

#### Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article:

**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 for performing meta-analyses that include such RCTs. Researchers are encouraged to report ICCs for neonatal outcomes in twins in their own RCTs.

### **Tweetable abstract**

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

### **Keywords**

Bayesian analysis; intraclass correlation coefficient; meta-analysis; power; sample size; twins

### **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.<sup>1</sup> Compared with singleton pregnancies, twins have a higher risk of adverse neonatal outcomes including preterm birth, respiratory distress syndrome, low birthweight, and mortality.<sup>2,3</sup> 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.<sup>4–9</sup>

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.<sup>10,11</sup> 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,<sup>12</sup> and in the analysis to avoid producing results that are over-precise.<sup>13</sup> 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 1 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. As the correlation between neonatal outcomes of twins is rarely reported in trial publications,<sup>14</sup> appropriately designing and analysing RCTs in women with twin pregnancies can be difficult and published estimates are needed.

The purpose of this study was 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

### Data sets

Twelve data sets including a total of 4504 twin pairs were used to estimate intraclass correlation coefficients (ICCs), as summarised in Tables S1 and S2. The data sets 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 this study. The first data set comes from a multicentre, open-label RCT assessing the effectiveness of a cervical pessary compared with no intervention for preventing poor perinatal outcomes.<sup>15</sup> The trial recruited 813 women with a multiple pregnancy between 12 and 20 weeks of 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 praevia. Women assigned to the cervical pessary group had a pessary inserted between 16 and 20 weeks of gestation and removed in the 36th week of gestation; women in the control group received standard antenatal care. Approximately 55% of women delivered preterm (less than 37 weeks of gestation).

The remaining data sets 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.<sup>16</sup> 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.<sup>4-9,17-21</sup> Inclusion/exclusion criteria and treatment regimens varied between these trials (Table S1). The study size ranged from 67 to 677 twin pairs, with trials either including both monozygotic and dizygotic twin pregnancies,<sup>4,5,7,17,18,21</sup> dizygotic twin pregnancies only<sup>6,8,19</sup> or not recording chorionicity<sup>9,20</sup> (Table S2). Preterm birth rates (<37 weeks of gestation) ranged from 50 to 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 (<2500 g and <1500 g); and two composite measures of adverse neonatal outcome, as defined in a previous study.<sup>16</sup> The first composite outcome included perinatal death, RDS, BPD, IVH, NEC, and sepsis, and 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 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 member 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, as ICCs calculated ignoring potential treatment effects may be biased,<sup>22</sup> 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.<sup>14,23</sup> 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<sup>24</sup> 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.<sup>25,26</sup> ICCs and 95% CIs 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.<sup>27</sup>

## 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 (Tables S3–S8). Mixed-effects models generally produced similar ICC estimates (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 at between 16 and 20 weeks of gestation to receive the new drug or placebo in a ratio of 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. Secondly, the sample size is multiplied by a quantity known as the design effect, which is given by  $1 + \text{ICC}$  for trials randomising and treating pregnant women and only including twin pregnancies.<sup>28</sup> 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  $1.68 \times 986 = 1658$  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 (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 to 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 might 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; 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  $1 + ICC$ .<sup>28</sup> 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. Our ICC estimates were generally above 0.5, indicating that RCTs focusing on twins are likely to require at least 50% more infants than are 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. However, this does not necessarily mean that appropriately powered RCTs in twins will be more expensive than trials in singletons, 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.<sup>28</sup>

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 GEEs or mixed-effects models.<sup>10,11,29–32</sup> 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.<sup>33</sup> 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<sup>34</sup> 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 two 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 than for 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 in 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.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## 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.

### 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.

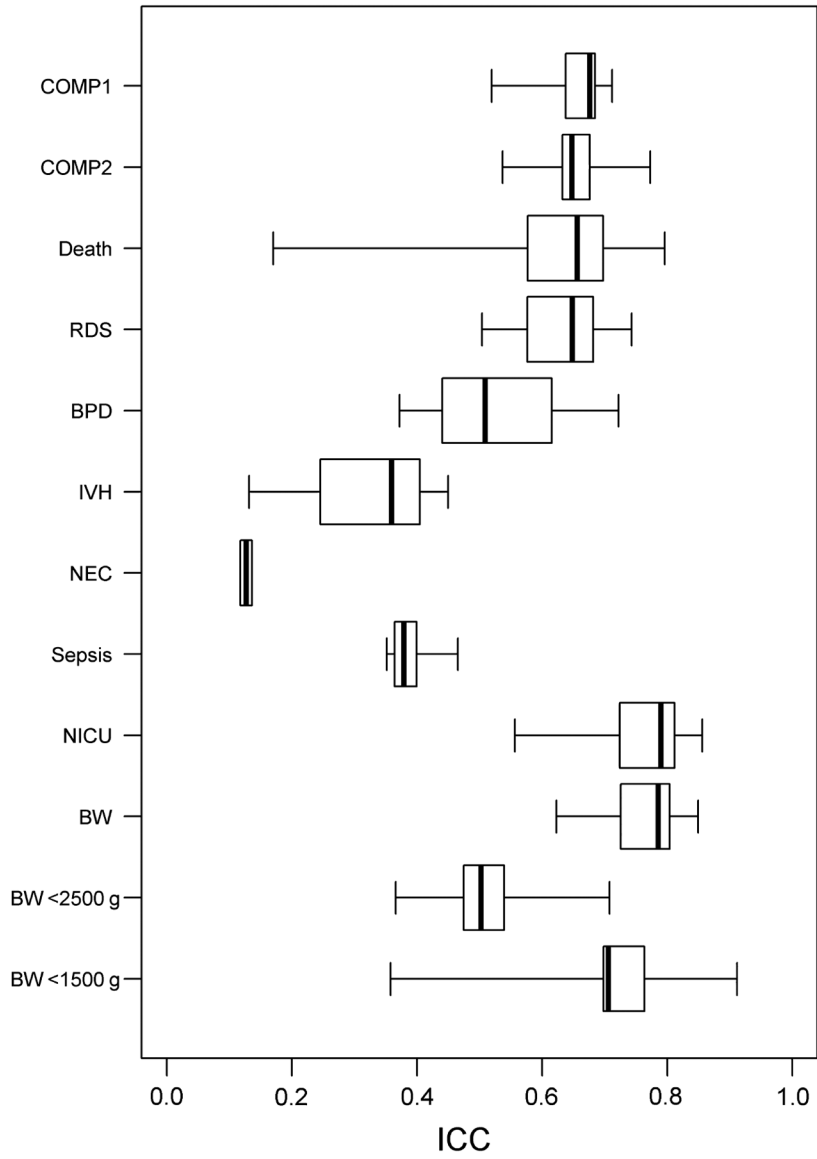
## References

1. Ananth CV, Chauhan SP. Epidemiology of twinning in developed countries. *Semin Perinatol* 2012;36:156–61. [PubMed: 22713495]



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. *Am J Public Health* 2002;92:1323–30. [PubMed: 12144992]
3. Shinwell ES, Haklai T, Eventov-Friedman S. Outcomes of multiples. *Neonatology* 2009;95:6–14. [PubMed: 18832859]
4. Rode L, Klein K, Nicolaidis KH, Krampfl-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 Obstet Gynecol* 2011;38:272–80. [PubMed: 21739497]
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;373:2034–40. [PubMed: 19523680]
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* 2013;120:50–7. [PubMed: 22882759]
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* 2015;122:71–9. [PubMed: 25163819]
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. *Am J Obstet Gynecol* 2011;204:221.e1–8. [PubMed: 21376161]
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. *Am J Obstet Gynecol* 2013;208:194.e1–8. [PubMed: 23433324]
10. Gates S, Brocklehurst P. How should randomised trials including multiple pregnancies be analysed? *BJOG* 2004;111:213–19. [PubMed: 14961881]
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. *Paediatr Perinat Epidemiol* 2009;23:380–92. [PubMed: 19523085]
12. Killip S, Mahfoud Z, Pearce K. What is an intracluster correlation coefficient? Crucial concepts for primary care researchers. *Ann Fam Med* 2004;2:204–8. [PubMed: 15209195]
13. Campbell MK, Piaggio G, Elbourne DR, Altman DG, Group C. Consort 2010 statement: extension to cluster randomised trials. *BMJ* 2012;345:e5661. [PubMed: 22951546]
14. Yelland LN, Sullivan TR, Makrides M. Accounting for multiple births in randomised trials: a systematic review. *Arch Dis Child Educ Pract Ed* 2015;100: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;382:1341–9. [PubMed: 23924878]
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* 2015;122:27–37. [PubMed: 25145491]
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. *N Engl J Med* 2007;357:454–61. [PubMed: 17671253]
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. *Obstet Gynecol* 2011;118:513–20. [PubMed: 21860279]
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. *Reprod Biomed Online* 2012;25:133–8. [PubMed: 22695310]

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. *J Perinat Med* 2012;40:593–9. [PubMed: 23093256]
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. *Arch Gynecol Obstet* 2011;283:423–9. [PubMed: 20091317]
22. Giraudeau B. Model mis-specification and overestimation of the intraclass correlation coefficient in cluster randomized trials. *Stat Med* 2006;25:957–64. [PubMed: 15977300]
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. *J Pediatr* 2010;156:202–8. [PubMed: 19969305]
24. Carpenter J, Bithell J. Bootstrap confidence intervals: when, which, what? A practical guide for medical statisticians. *Stat Med* 2000;19:1141–64. [PubMed: 10797513]
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;63:935–41. [PubMed: 17825023]
26. Mancl LA, DeRouen TA. A covariance estimator for GEE with improved small-sample properties. *Biometrics* 2001;57:126–34. [PubMed: 11252587]
27. SAS Institute Inc. jackboot.sas [<http://support.sas.com/kb/24/982.html>] Accessed 3 July 2008.
28. Yelland LN, Sullivan TR, Price DJ, Lee KJ. Sample size calculations for randomised trials including both independent and paired data. *Stat Med* 2017;36:1227–39. [PubMed: 28074483]
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. *Ann Epidemiol* 2005;15:293–301. [PubMed: 15780777]
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;17:110. [PubMed: 28728549]
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. *Stat Med* 2013;32:1429–38. [PubMed: 23027676]
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? *Paediatr Perinat Epidemiol* 2011;25:283–97. [PubMed: 21470268]
33. Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB. *Bayesian Data Analysis*, 3rd edn. Boca Raton, FL: CRC Press, 2014.
34. Higgins JPT, Green S (editors). *Cochrane Handbook for Systematic Reviews of Interventions*. The Cochrane Collaboration, 2011. [www.handbook.cochrane.org](http://www.handbook.cochrane.org).



**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, necrotising enterocolitis; NICU, neonatal intensive care unit; BW, birthweight; ICC, intraclass correlation coefficient.

**Table 1.**

Summary of intraclass correlation coefficient (ICC) estimates for neonatal outcomes across trials

Outcome	Median (range) ICC	Trials
Composite adverse neonatal outcome 1 *	0.68 (0.52–0.71)	5–9,15,17,18,21
Composite adverse neonatal outcome 2 **	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 <2500 g	0.50 (0.37–0.71)	4–9,15,17–21
Birthweight <1500 g	0.71 (0.36–0.91)	4–9,15,17,18,20,21

\* Includes perinatal death, respiratory distress syndrome, bronchopulmonary dysplasia, intraventricular haemorrhage, necrotising enterocolitis, and sepsis.

\*\* Includes perinatal death, respiratory distress syndrome, intraventricular haemorrhage, and necrotising enterocolitis.