










Good practices for the design, analysis, and interpretation of observational studies on birth spacing and perinatal health outcomes

Jennifer A. Hutcheon¹  | Susan Moskosky² | Cande V. Ananth^{3,4}  |
 Olga Basso^{5,6}  | Peter A. Briss⁷ | Cynthia D. Ferré⁸ | Brittni N. Frederiksen² |
 Sam Harper⁶ | Sonia Hernández-Díaz⁹ | Ashley H. Hirai¹⁰ | Russell S. Kirby¹¹  |
 Mark A. Klebanoff¹²  | Laura Lindberg¹³ | Sunni L. Mumford¹⁴  | Heidi D. Nelson¹⁵ |
 Robert W. Platt⁶ | Lauren M. Rossen¹⁶ | Alison M. Stuebe^{17,18}  | Marie E. Thoma¹⁹  |
 Catherine J. Vladutiu¹⁰ | Katherine A. Ahrens² 

¹Department of Obstetrics and Gynaecology, University of British Columbia, Vancouver, British Columbia, Canada

²Office of Population Affairs, Office of the Assistant Secretary for Health, Rockville, Maryland

³Department of Obstetrics and Gynecology, Irving College of Physicians and Surgeons, Columbia University, New York, New York

⁴Department of Epidemiology, Joseph L. Mailman School of Public Health, Columbia University, New York, New York

⁵Department of Obstetrics and Gynecology, Royal Victoria Hospital, Research Institute of McGill University Health Centre, Montreal, Quebec, Canada

⁶Department of Epidemiology, Biostatistics, and Occupational Health, McGill University, Montreal, Quebec, Canada

⁷National Center for Chronic Disease Prevention and Health Promotion, Centers for Disease Control and Prevention, Atlanta, Georgia

⁸Maternal and Infant Health Branch, Division of Reproductive Health, National Center for Chronic Disease Prevention and Health Promotion, Centers for Disease Control and Prevention, Atlanta, Georgia

⁹Department of Epidemiology, Harvard T. H. Chan School of Public Health, Boston, Massachusetts

¹⁰US Department of Health and Human Services, Health Resources and Services Administration, Maternal and Child Health Bureau, Office of Epidemiology and Research, Rockville, Maryland

¹¹Department of Community and Family Health, University of South Florida College of Public Health, Tampa, Florida

¹²Division of Epidemiology, Departments of Pediatrics and Obstetrics and Gynecology, Center for Perinatal Research, The Research Institute at Nationwide Children's Hospital, The Ohio State University, Columbus Ohio

¹³Guttmacher Institute, New York, New York

¹⁴Epidemiology Branch, Division of Intramural Population Health Research, National Institute of Child Health and Human Development, Bethesda, Maryland

¹⁵Department of Medical Informatics and Clinical Epidemiology, Oregon Health and Science University, Portland, Oregon

¹⁶Reproductive Statistics Branch, Division of Vital Statistics, Centers for Disease Control and Prevention, National Center for Health Statistics, Hyattsville, Maryland

¹⁷Department of Obstetrics and Gynecology, University of North Carolina School of Medicine, Chapel Hill, North Carolina

¹⁸Department of Maternal and Child Health, Gillings School of Global Public Health, Chapel Hill, North Carolina

¹⁹Department of Family Science, University of Maryland, College Park, Maryland

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2018 The Authors. *Paediatric and Perinatal Epidemiology* Published by John Wiley & Sons Ltd.



Correspondence

Jennifer Hutcheon, BC Children's and Women's Hospital, Vancouver, BC, Canada.
Email: jhutcheon@bcchr.ca

Funding information

This manuscript was supported, in part, by a contract between the Office of Population Affairs and Atlas Research, LLC [# HHSP233201450040A]. JAH is the recipient of New Investigator Awards from the Canadian Institutes of Health Research and the Michael Smith Foundation for Health Research. SLM was supported by the Intramural Research Program of the Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institutes of Health.

Abstract

Background: Meta-analyses of observational studies have shown that women with a shorter interpregnancy interval (the time from delivery to start of a subsequent pregnancy) are more likely to experience adverse pregnancy outcomes, such as preterm delivery or small for gestational age birth, than women who space their births further apart. However, the studies used to inform these estimates have methodological shortcomings.

Methods: In this commentary, we summarise the discussions of an expert workgroup describing good practices for the design, analysis, and interpretation of observational studies of interpregnancy interval and adverse perinatal health outcomes.

Results: We argue that inferences drawn from research in this field will be improved by careful attention to elements such as: (a) refining the research question to clarify whether the goal is to estimate a causal effect vs describe patterns of association; (b) using directed acyclic graphs to represent potential causal networks and guide the analytic plan of studies seeking to estimate causal effects; (c) assessing how miscarriages and pregnancy terminations may have influenced interpregnancy interval classifications; (d) specifying how key factors such as previous pregnancy loss, pregnancy intention, and maternal socio-economic position will be considered; and (e) examining if the association between interpregnancy interval and perinatal outcome differs by factors such as maternal age.

Conclusion: This commentary outlines the discussions of this recent expert workgroup, and describes several suggested principles for study design and analysis that could mitigate many potential sources of bias.

KEYWORDS

adverse perinatal outcomes, birth spacing, causal inference, epidemiologic bias, interpregnancy interval, preterm birth

1 | INTRODUCTION

Interpregnancy interval, the time from delivery until start of a subsequent pregnancy, has consistently been associated with perinatal health outcomes.¹ Meta-analysis has shown that interpregnancy intervals of less than 6 months are associated with a 40% increased odds of delivery before 37 weeks' gestation, and 26% increased odds of small for gestational age birth compared with intervals of 18-23 months.¹ Interpregnancy intervals of 6-11 and 12-17 months are also associated with increased risks for these outcomes. As a result, the proportion of births delivered after an interpregnancy interval of less than 18 months is monitored as a Healthy People 2020 Indicator,² and the American College of Obstetricians and Gynecologists currently states that "the optimal interval between delivery and subsequent pregnancy is 18 months to 5 years."³

Much of the literature on interpregnancy interval comes from low- to middle-income countries, where women's baseline nutritional status, health, and health care differs substantially from women in the United States (US).¹ It is unclear whether the associations

between short interpregnancy interval and adverse health outcomes, which are believed to be mediated, at least partly, through nutritional depletion⁴ are comparable between these populations. In September 2017, the US Department of Health and Human Services Office of Population Affairs convened an expert workgroup to review available evidence on birth spacing (ie, interpregnancy interval) and pregnancy outcomes within high-resource settings. The expert workgroup recognised the inherent challenges of using observational data to disentangle the effects of interpregnancy interval from underlying socio-economic factors, pregnancy intention, and access to health services, but nevertheless agreed that the inferences drawn from observational studies could be improved by addressing a number of other methodological shortcomings.

The goal of this review is to outline good practices identified by the expert workgroup for the design, analysis, and interpretation of observational studies of interpregnancy interval and perinatal outcomes, with the aim of improving the quality of evidence available for public health recommendations on birth spacing in the United States. We first discuss good practices for all observational studies

on interpregnancy intervals, and then discuss additional considerations for studies employing a sibling comparison design. We focus primarily on research studies that investigate the causal effect of short interpregnancy interval following a live birth on adverse perinatal outcomes. Our good practices outline builds upon the more general guidelines for reporting observational studies that have been published and updated (STROBE statement).⁵ The recommendations contained herein are based upon expert opinion and general epidemiological literature; they are not based on new empirical data from analyses we conducted on interpregnancy intervals and adverse pregnancy outcomes.

This manuscript is part of a theme issue on the Office of Population Affairs' expert workgroup meeting on birth spacing and adverse pregnancy outcomes, which includes a separate summary of the overall meeting proceedings.^{6–10}

2 | RESEARCH OBJECTIVE

2.1 | Clearly specify the research question

There are multiple questions of interest within the general topic of interpregnancy interval and adverse birth outcomes. Typically, researchers are interested in isolating the physiological effect of interpregnancy interval on health outcomes (eg, “does a short interpregnancy interval *cause* an adverse birth outcome?”). However, researchers may also seek to establish whether interpregnancy interval is a predictor of adverse birth outcomes to help identify higher-risk patients in antenatal care (eg, “is short interpregnancy interval *predictive* of adverse birth outcomes?”).

Being clear about the study's specific research question is important because the analytic approach needed to answer each question differs. For example, if the goal is to isolate the total causal effect of interpregnancy interval on adverse perinatal outcomes, researchers should ensure that they have controlled for all other “open backdoor” pathways between interpregnancy interval and adverse outcomes created by confounders,¹¹ such as socioeconomic position and pregnancy intention. It is also advisable to consider the target trial that the observational study seeks to emulate (eg, eligibility criteria, treatment definition, follow-up, outcome, causal contrast).¹² For predictive questions, however, the timing of when information for each variable becomes available in the real world setting is critical.¹³ For example, if the goal is to identify whether interpregnancy interval can be used to predict

women at increased risk of adverse pregnancy outcomes, such as preterm birth at the time of a first prenatal visit, it is problematic to include variables that only become available at the time of delivery (such as total pregnancy weight gain or infant birthweight). In addition, factors that are not confounders but are highly predictive should be included in the predictive model. Misaligning the analytic plan with the specific research question may result in study estimates that fail to answer the question of interest.

2.2 | If the goal is to estimate a causal effect, the hypothesised relationships between study variables should be explicitly described

Directed acyclic graphs (DAG) are recommended as a conceptual framework to delineate these relationships and inform the choice of variables to include in adjusted analyses.^{14,15} In addition, DAGs show how bias can be introduced by controlling (through adjustment, matching, stratification, or restriction) for variables that are common mediators between interpregnancy interval and adverse pregnancy outcomes. For example, in a study seeking to evaluate the effect of interpregnancy interval on the risk of subsequent preterm birth, adjustment for prepregnancy body mass index (BMI) at the start of the subsequent pregnancy may be inappropriate (illustrated in Figure 1). This is because women with a short interpregnancy interval have had less time to lose their pregnancy weight postpartum; thus higher prepregnancy BMI at the start of the subsequent pregnancy may be in the causal pathway between short interpregnancy interval and preterm birth. Mapping out hypothesised relationships can also inform data collection and highlight potential concerns regarding confounding from unmeasured variables.

3 | STUDY DESIGN

3.1 | Consider quasi-experimental designs

The expert workgroup recognised that conventional observational studies with adjustment for measured confounders may fall short of fully eliminating confounding arising from underlying socioeconomic factors and differences in access to health care services. Researchers should consider quasi-experimental designs, such as interrupted time series and difference-in-differences analyses, as alternative strategies to control for confounding.¹⁶ The rationale behind quasi-experimental designs is that changes in policies or clinical practice can sometimes create the structure of a natural experiment

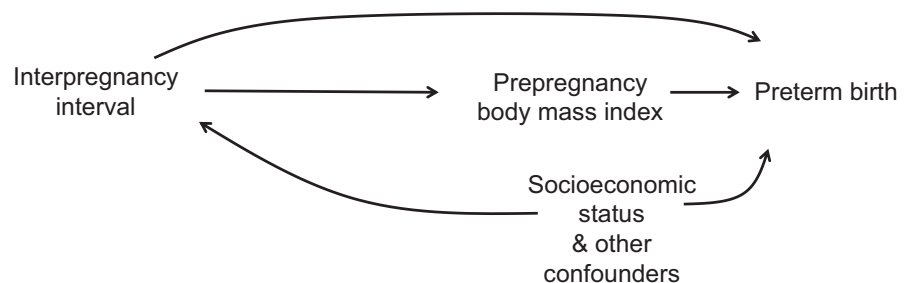


FIGURE 1 Simplified example of using directed acyclic graph as a conceptual framework to delineate the relationship between interpregnancy interval and preterm birth, with a mediating path through prepregnancy body mass index and potential confounders shown



where exposure to an intervention approximates random assignment.¹⁷ Because the introduction of a new policy or change in practice is typically unrelated to women's individual characteristics, careful comparisons of those affected and unaffected by the policy change or change in practice can produce estimates less prone to confounding by individual-level characteristics.¹⁶

Researchers have used quasi-experimental designs based on state-level differences in the diffusion of oral contraceptives in the 1960s and 1970s to understand how the ability to prevent pregnancy affects a woman's longer-term economic earnings.¹⁸ Likewise, policy or practice changes could be used to investigate the effect of modifying interpregnancy intervals on adverse pregnancy outcomes, such as through broader access to paid parental leave, subsidised childcare, family housing, and education, or financial incentives for having children, like tax deductions—all of which could encourage short birth spacing. Natural experiments that increase birth spacing often occur through increased access to contraception or abortion which, in turn, will reduce births resulting from unintended pregnancies. For these studies, researchers should be cautious about attributing any differences in perinatal outcome to increases in interpregnancy interval independent of a reduction in unintended pregnancy. Natural experiments that act primarily through mechanisms other than provision or restriction of contraception and abortion services (eg, social policies mentioned above or external events such as spousal deployment in military families) may be most valuable in isolating the causal effect of interpregnancy interval.

4 | CHARACTERISING THE INTERPREGNANCY INTERVAL

4.1 | Use interpregnancy interval (birth to start of next pregnancy), rather than interbirth interval (birth to birth)

Interbirth interval is determined in part by the duration of a woman's subsequent pregnancy. As pregnancies that result in adverse perinatal outcomes (such as stillbirth or preterm birth) are often shorter than those resulting in healthy deliveries,^{19,20} there is a direct correlation between short interbirth interval and adverse outcomes that is independent of any causal effect of interpregnancy interval. This potential for bias does not occur when using the measure of interpregnancy interval. Further, interpregnancy interval better approximates the modifiable health behaviour of interest, the duration of time after delivery that women at risk of pregnancy use contraception correctly and consistently to prevent pregnancy.

4.2 | Account for non-linear associations

Both short and long interpregnancy intervals have been linked with increased risks of adverse perinatal outcomes.¹ Modelling interpregnancy interval as a continuous linear variable in multivariable regression will not capture this pattern and will produce a biased estimate of the association between short interpregnancy interval

and adverse outcome that is attenuated towards the null.²⁴ Instead, this reverse J-shape association should be analysed using non-parametric smoothing methods such as restricted cubic splines or fractional polynomials.^{25,26} Calculating risk ratios and/or risk differences for select intervals along the continuum of interpregnancy interval (ie, at 6, 12, 18 months) may be helpful for interpretation.

A simpler approach is to create categorical variables that capture discrete segments of the interpregnancy interval. While retaining interpregnancy interval as a continuous variable will reduce loss of information,²⁷ categorisation will facilitate comparison with previous studies as well as with existing recommendations. Using interpregnancy interval categories of <6, 6-11, 12-17, 18-23, 24-59, and ≥60 months, with 18-23 as a reference category, is recommended.^{1,28} Given the advantages of each approach, presenting findings for interpregnancy interval in both continuous and categorical form is worthwhile.

4.3 | Assess how miscarriages and pregnancy terminations may have influenced interpregnancy interval classifications

Pregnancy losses prior to 20 weeks' gestation (eg, miscarriages and pregnancy terminations) are not reliably captured in population registries or individual-level surveys used to study interpregnancy interval.²⁹ In the absence of this information, interpregnancy interval is calculated as the time from delivery until start of the subsequent pregnancy lasting past 20 weeks' gestation, not necessarily the subsequent pregnancy.³⁰ For a woman with an intervening pregnancy loss, her interpregnancy interval will be longer, as it incorporates the time from loss to when the woman became pregnant again,³¹ and inclusion of these women may no longer reflect the target population for whom advice on interpregnancy interval planning is intended.¹⁹

Sensitivity analyses should be undertaken to assess the potential influence of intervening early pregnancy loss on the interpregnancy interval-adverse outcome association. Ideally, analyses should be repeated in a cohort with presumed accurate classification of interpregnancy interval (ie, women with no intervening miscarriages or terminations). As such detailed data are rarely available, this cohort should be recreated as closely as possible using additional available data such as gravidity (the number of times a woman has been pregnant); for example, a cohort of women whose gravidity matches parity. However, early miscarriages may be undetected by a woman, and still be correlated with other poor pregnancy outcomes. Researchers should highlight these limitations of their analyses and moderate their interpretation of results accordingly.

5 | ACCOUNTING FOR COVARIATES

5.1 | Consider the potential for effect measure modification

The expert workgroup identified several factors that could potentially modify the association between interpregnancy interval

and adverse health outcomes. For example, it was hypothesised that the risks associated with short interpregnancy interval may be greater in younger women compared with older women. Other potential effect measure modifiers include previous caesarean delivery, socio-economic position, immigrant status (eg, foreign born vs US born), previous perinatal losses, race/ethnicity, and prepregnancy BMI at the initial pregnancy. Researchers should consider the potential for effect measure modification a priori and address this by presenting stratified analyses or including a product term between interpregnancy interval and the potential effect measure modifier in a multivariable model. It is preferable to evaluate deviations from joint effects on an additive scale.³² These deviations can identify subpopulations most likely to benefit from interventions to lengthen interpregnancy interval, and therefore are of greatest interest to policy-makers. Given the large sample sizes obtained from data sources such as US birth certificate records, researchers should avoid relying solely on statistical significance to make inferences about departures from additive effects; rather, they should also examine the clinical or public health importance of any differences.

5.2 | Address the roles of maternal socio-economic position, previous perinatal loss, and pregnancy intention

As in any epidemiologic study, researchers should ensure that they identify and account for all key covariates that could confound the association between interpregnancy interval and adverse perinatal outcomes. In doing so, researchers should ensure that they specifically consider the role of three key causes of adverse perinatal outcomes known to be strongly associated with short interpregnancy interval: maternal socio-economic position, previous perinatal loss, and pregnancy intention.

Women of lower socio-economic position are considerably more likely to have shorter interpregnancy intervals, and this position is also a risk factor for adverse pregnancy outcomes.³³ This creates a potential for confounding and effect measure modification (as discussed above).

Women experiencing a perinatal loss (neonatal death or still-birth) are more likely to subsequently have a short interpregnancy interval than women with a live infant.³⁴ If there is an underlying condition causing adverse outcomes in both pregnancies, this could create a spurious link between short interval and adverse outcome in the subsequent pregnancy. Depending on the research question, it may be preferable to restrict the study cohort to women whose pregnancy preceding the interpregnancy interval did not end in still-birth or neonatal death. Previous methodological work on the role of past reproductive history may be useful in establishing the best approach to account for past perinatal loss.^{35,36}

Pregnancies following short interpregnancy intervals are disproportionately more likely to represent unintended pregnancies,³⁷ which have previously been linked with increased risks of adverse perinatal outcomes.³⁸ Although pregnancy intention is a complex

construct that is challenging to measure, and may be closely interwoven with socio-economic position and culture, efforts should be made to ensure that consequences of unintended pregnancy on perinatal health are not incorrectly attributed to a causal effect of short interpregnancy interval.

Many data sources, such as US Natality files, will not have measures of each of these above factors or will only measure women's socio-economic status at the time of the interview, limiting the value of these data sources to adequately isolate the effects of interpregnancy interval.

5.3 | Be clear about the timing of measurement for each variable

Values for many covariates such as maternal age, smoking status, prepregnancy BMI, and co-morbidity status (gestational diabetes, preeclampsia) may differ between the time of an initial and subsequent pregnancy. When establishing and reporting analytic plans, researchers should specify at which time point a variable was measured. Clarifying when a variable was measured is essential for determining whether the appropriate adjustment strategy was used. A useful exercise to determine the most appropriate covariates to include in the analysis, based on the timing of their occurrence, is to imagine how the research question would be answered using a hypothetical randomised clinical trial.³⁹ In such a trial, randomisation would be expected, on average, to produce two groups that are balanced with respect to their baseline characteristics at the time of the intervention. For interpregnancy interval, a randomised design would likely involve randomisation immediately postpartum of women to different interpregnancy intervals, with follow-up of perinatal outcomes in the subsequent pregnancy should it occur. In an observational study, the balancing of study groups would be achieved by confounder adjustment, such as through multivariable regression (or other methods such as propensity score analysis and latent variable analysis), rather than randomisation. As a result, covariates measured at the time of or prior to the initial delivery, rather than during the interpregnancy interval or subsequent pregnancy, will be the most appropriate for inclusion in adjusted analyses. This is particularly important for maternal age, as outlined below. As a general principle, it is beneficial to address these issues at the design stage rather than at the time of statistical analysis.

5.4 | Be thoughtful about adjusting for maternal age at second pregnancy

Maternal age at subsequent pregnancy is the direct result of a woman's age at the initial delivery and the length of the interpregnancy interval. For example, adolescent mothers at the time of their second pregnancy will necessarily have had a short interpregnancy interval in order to have had their second child while still under 20 years of age. A study on interpregnancy interval restricted to adolescent mothers giving birth to their second child will, by design, have been enriched for births preceded by short interpregnancy intervals. For

most research questions, only maternal age at the initial pregnancy should be controlled for (ideally at the time of delivery), as this better reflects the time point at which differences in age distributions between groups should be made comparable. For studies using the publicly-released data of the US Natality files, for instance, this may require access to restricted variables (such as month and year of mother's birth, rather than age in years) to more precisely calculate a woman's age at her previous pregnancy.

Interpregnancy interval, maternal age at initial delivery, and maternal age at subsequent delivery are multi-collinear and should never be included simultaneously in a regression model. If included as continuous variables, they will be perfectly collinear and one variable will be inestimable. While categorisation or rounding of these variables may allow all three to be added simultaneously, estimates will be highly unstable. Additionally, models will be incorrectly specified as only maternal age at index delivery meets the definition of a confounder because it precedes the interpregnancy interval.^{40,41}

6 | REPORTING RESULTS

6.1 | Present results in both absolute and relative terms

Reporting the risk associated with an exposure exclusively in relative terms (eg, using measures of effect such as the odds ratio or risk ratio) can lead patients, providers, and policy-makers to overestimate its harm (or benefit).^{42,43} Studies of interpregnancy interval should, if possible, report the absolute risks of adverse outcomes between different interpregnancy interval categories using risk differences (ie, excess number of cases per 100 deliveries) in addition to or instead of relative measures. Absolute probabilities may be

more useful for patient counselling on the consequences of short interpregnancy interval. Adjusted risks and risks differences with 95% confidence intervals can be readily calculated using the "margins" command in R or Stata; in SAS, a macro may be needed.³⁸

6.2 | Exercise caution in interpretation

Public health recommendations on birth spacing reflect a balancing of risks associated with both short and long interpregnancy intervals. However, studies of interpregnancy interval and adverse outcomes, conducted among women with two or more births, miss an important consequence of delaying a subsequent pregnancy: that the rise in age-related infertility may prevent the occurrence of a second birth.⁴⁴ Thus, researchers should be cautious about making statements about optimal intervals based solely on their findings in a study associating interpregnancy interval with adverse outcomes among women who had two or more births.

7 | SIBLING COMPARISON STUDIES

The expert workgroup discussed study design issues specifically relevant to sibling comparison studies, that is studies that use a woman as her own control by comparing the interpregnancy intervals and perinatal outcomes of a woman's different pregnancies (siblings). As illustrated in Figure 2, sibling comparison studies of interpregnancy interval are necessarily restricted to women with three or more pregnancies (ie, two or more interpregnancy intervals). The design takes advantage of the fact that siblings are typically exposed to similar family conditions (genetic or environmental), and enables researchers to eliminate the effects

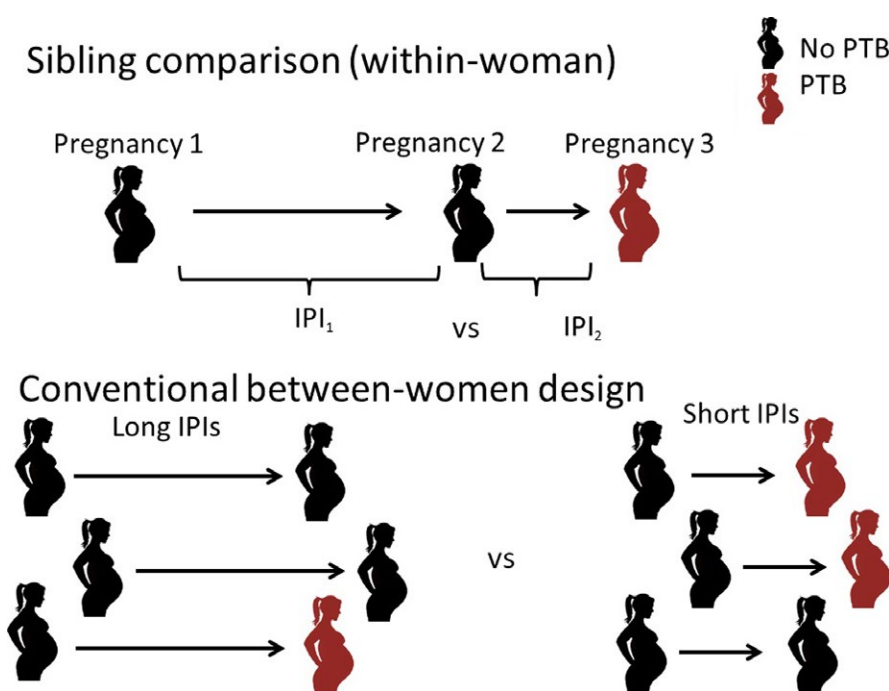


FIGURE 2 Schematic contrasting the sibling comparison design (within-woman) with the conventional between-women cohort design in the study of interpregnancy interval (IPI)

of any confounding characteristics that remain constant within a woman, but are difficult to measure or control for using standard regression approaches (such as some aspects of socio-economic position or family disease history).^{45,46} Despite enthusiasm for the design's ability to control for difficult to measure confounders, the expert workgroup agreed that the credibility of assumptions underlying this design and generalisability of results require more investigation.

7.1 | Assess the generalisability of the cohort used in the sibling analysis

The sibling comparison design analyses the study cohort as a series of matched pairs, where typically siblings from a woman's second and third pregnancy (ie, the births following the first and second interpregnancy interval, respectively) are compared to each other. As with any matched-pair design, only pairs that are discordant with respect to exposure and outcome contribute to the overall estimate of effect.⁴⁷ Most adverse perinatal outcomes are relatively rare, so only a small fraction of siblings will be discordant with respect to the study outcome. For example, an outcome with a risk of approximately 10% (such as preterm birth or small for gestational age birth) will, by definition, theoretically have a maximum of 20% of women experiencing discordant outcomes in their two pregnancies. As some women experience adverse outcomes in both pregnancies, this percentage will be even lower. The cohort of siblings used to generate estimates of the exposure-outcome association is therefore markedly smaller than the entire cohort of eligible sibling pairs (ie, siblings born to women with three or more births), which, in turn, is smaller than the population of women to whom future birth spacing recommendations are intended (women with at least one live birth who may become pregnant again).

Given the restricted nature of the analytic cohort used in sibling analyses, researchers should present data to assess its generalisability. First, they could compare the characteristics of women with discordant perinatal outcomes in their second and third pregnancies to those with concordant outcomes. Second, they could compare results of the sibling comparison design (often obtained from conditional logistic regression) to estimates derived from the conventional between-women analysis (often obtained from unconditional logistic regression) restricted to the analytic cohort providing informative data for the sibling comparison analyses (ie, women with discordant birth outcomes).

A direct comparison of the effect estimate, such as an odds ratio, of an association between interpregnancy interval and an adverse perinatal outcome between a sibling design vs a cohort design should be avoided.^{48,49} A sibling design analysed using a conditional logistic regression model is regarded as a "subject-specific" model. In contrast, an association estimated from a cohort analysis depicts a model of the "marginal means," producing a population-averaged estimate. Odds ratios estimated from a subject-specific regression approach will be smaller, on average, than one estimated from a population-averaged model.⁵⁰ The interpretation of effect measures

from a subject-specific and population-averaged models differ, and consequently, do not enable a direct head-to-head comparison, and, as previously discussed, the study populations used to derive the estimates differ.

7.2 | Keep exposures and outcomes continuous when possible

As described in the previous section, sibling comparison analyses are only informed by sibling pairs with discordant exposure and outcome status. Categorisation of continuous variables reduces the number of response options, and thus increases the likelihood of concordant exposure or outcome status between pregnancies. In addition, categorisations may lead to compromised statistical power if the category boundaries are not optimal, and could introduce selection bias by selecting for discordant pairs.⁵¹ Where possible, continuous variables, such as interpregnancy interval and birthweight for gestational age percentiles, should remain in continuous form in sibling comparison studies.

7.3 | Adjust for confounders that vary between pregnancies

The sibling comparison design controls for confounders that remain constant across a woman's pregnancies (eg, race/ethnicity), but does not control for characteristics that vary across a woman's pregnancies.⁵² Time-varying characteristics that could potentially differ across pregnancies include pregnancy intention, smoking status, infections, relationship status, social support, living conditions, income, maternal co-morbidity status, and weight. As previously reported,⁵¹ uncontrolled time-varying confounding in the sibling comparison design can introduce a greater degree of bias than in conventional observational studies. As a result, researchers should identify and adjust for relevant time-varying confounders when using the sibling comparison design.

The extent to which the sibling comparison design magnifies bias due to uncontrolled non-shared confounders depends on the correlation between siblings in the exposure of interest (eg, interpregnancy interval) relative to the confounders. Bias will be greater if the between-sibling correlation in interpregnancy interval is stronger than the between-sibling correlation in confounders.⁵¹ Calculating and reporting these correlations are recommended to help assess the likelihood of magnified bias by other, unmeasured confounders.

7.4 | Examine potential carry-over effects

The validity of conclusions drawn from a sibling comparison design rests on an assumption that the pregnancy conditions and outcomes following a woman's first interpregnancy interval are wholly independent from those of pregnancies following a subsequent interpregnancy interval, conditional on measured confounders. In addition to the pregnancies having to be exchangeable with respect to covariates, there can also be no carry-over effect of the exposure



itself.⁴⁷ The extent to which this assumption holds for sibling comparison studies of interpregnancy intervals is unclear,⁵³ and it is worth noting that the design was intended primarily for transient or acute exposures. One could speculate that any effects of close pregnancy spacing mediated through nutritional status could differ if there was a cumulative effect of nutritional depletion over three, rather than two, pregnancies.⁴ Likewise, one could speculate that the impact of a short interpregnancy interval on child injury may differ if a family has two other young children at home competing for attention, rather than only one.^{54,55}

Although it is difficult to prove the validity of this assumption, researchers can explore its plausibility in several ways. First, they can present results stratified by birth order to confirm that estimates remain similar irrespective of whether the short interpregnancy interval was a woman's first vs second interval. More formally, they can include a product term between interpregnancy interval and birth order to test for effect measure modification by birth order.

7.5 | Assess the validity of the positivity assumption

Valid estimates of the average causal effect of interpregnancy interval on health outcomes depend on an assumption that the condition of positivity is met.¹¹ Positivity means that there is a positive probability of observing every level of exposure for every combination of values of exposure and confounders. Although this assumption is required for any study aiming for causal inference, it seems particularly problematic for sibling comparison studies. For example, it is impossible to find two singletons from the same mother that have the same parity and maternal age at delivery, both potential confounders of the association between interpregnancy interval and adverse outcomes. Adjustments for such factors therefore guarantee positivity violations. The extent to which these violations impact estimates of effect is unclear. However, researchers should explore the magnitude of this problem by tabulating the number of observations in each combination of exposure and covariate grouping using a matched-pair table design. Creating a table that cross-tabulates each possible interpregnancy interval classification (eg, <6 months in the first interpregnancy interval and 18-23 months in the second interpregnancy interval) with levels of covariates will help to visualise the number of sparse or empty cells as an indicator of low probability data for each cell. Based on these findings, the extent to which the positivity assumption is likely credible for a given research question should be explicitly discussed.

8 | CONCLUSION

Despite the large body of research linking interpregnancy interval and adverse birth outcomes, a recent expert workgroup concluded that high-quality evidence to inform public health

recommendations on birth spacing is limited due to methodological limitations of existing studies. Observational research in this field poses numerous challenges, most notably controlling for confounding by socio-economic position and other factors. This commentary outlines the discussions of this recent expert workgroup, and describes several suggested principles for study design and analysis that could mitigate many potential sources of bias. These recommended principles will also be useful in understanding the strengths and limitations of new publications and for determining the risk of bias when selecting studies for statistical meta-analysis. Increasing the methodological rigour in the design and analysis of research on interpregnancy interval will greatly improve the inferences that can be drawn from this work, and ensure that future clinical and public health recommendations for birth spacing in the United States, and elsewhere, can be informed by the best possible estimates of the causal effects of interpregnancy interval on maternal and newborn health.

ACKNOWLEDGEMENT

The authors thank Jamie Hart and Julia Rollison, from Atlas Research, for facilitating the expert workgroup meeting Birth Spacing and Adverse Pregnancy Outcomes, in Washington, DC 14-15 September 2017. The authors also acknowledge the critical feedback they received during the meeting from the following participants: Maureen Norton, United States Agency for International Development; Lorrie Gavin, private consultant; Ann Borders, Northwestern University; and Karen Pazol, Centers for Disease Control and Prevention.










CONFLICT OF INTEREST AND FINANCIAL DISCLOSURES

Peter Briss, Lauren Rossen, and Cynthia Ferré work for the Centers for Disease Control and Prevention, an agency that published the Providing Quality Family Planning Services Recommendations with the Office of Population Affairs in 2014. Mark Klebanoff noted that his participation in the meeting was not intended to disqualify researchers working at The Ohio State University from responding to future requests for proposals from the Office of Population Affairs. No other financial or other disclosures of conflict of interest were reported by the authors of this paper at the time of the meeting or writing of this report.

DISCLAIMER

The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the Office of Population Affairs, Office of the Secretary for Health, Centers for Disease Control and Prevention, National Center for Health Statistics, Health Resources and Services Administration, or Maternal and Child Health Bureau.

ORCID

- Jennifer A. Hutcheon  <http://orcid.org/0000-0003-0502-3423>
- Cande V. Ananth  <http://orcid.org/0000-0002-0410-2595>
- Olga Basso  <http://orcid.org/0000-0001-9298-4921>
- Russell S. Kirby  <http://orcid.org/0000-0002-3489-401X>
- Mark A. Klebanoff  <http://orcid.org/0000-0003-0513-4975>
- Sunni L. Mumford  <http://orcid.org/0000-0003-0133-3176>
- Alison M. Stuebe  <http://orcid.org/0000-0002-1228-4587>
- Marie E. Thoma  <http://orcid.org/0000-0001-9267-4384>
- Katherine A. Ahrens  <http://orcid.org/0000-0001-5139-9208>

REFERENCES

- Conde-Agudelo A, Rosas-Bermudez A, Kafury-Goeta AC. Birth spacing and risk of adverse perinatal outcomes: a meta-analysis. *J Am Med Assoc.* 2006;295:1809-1823.
- US Department of Health and Human Services. Healthy People 2020 Objectives: Family Planning. Washington, DC. <http://www.healthypeople.gov/2020/topicsobjectives2020/overview.aspx?topicid=13>. Accessed October 17, 2013.
- Committee opinion No. 666: optimizing postpartum care. *Obstet Gynecol.* 2016;127:e187-e192.
- Conde-Agudelo A, Rosas-Bermudez A, Castano F, Norton MH. Effects of birth spacing on maternal, perinatal, infant, and child health: a systematic review of causal mechanisms. *Stud Fam Plann.* 2012;43:93-114.
- von Elm E, Altman DG, Egger M, Pocock SJ, Gotsche PC, Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: guidelines for reporting observational studies. *Int J Surg.* 2014;12:1495-1499.
- Ahrens K, Hutcheon J, Ananth C, et al. Report of the Office of Population Affairs' expert work group meeting on short birth spacing and adverse pregnancy outcomes: methodological quality of existing studies and future directions for research. *Paediatr Perinat Epidemiol.* 2018, In press.
- Ahrens KA, Nelson HD, Stidd R, Moskosky S, Hutcheon JA. Short interpregnancy intervals and adverse perinatal outcomes in high-resource settings: an updated systematic review. *Paediatr Perinat Epidemiol.* 2018, In press.
- Hutcheon JA, Nelson HD, Stidd R, Moskosky S, Ahrens KA. Short interpregnancy intervals and adverse maternal outcomes in high resource settings: an updated systematic review. *Paediatr Perinat Epidemiol.* 2018, In press.
- Thoma M, De Silva D, McDorman M. Examining interpregnancy intervals and maternal and perinatal health outcomes using U.S. vital records: important considerations for analysis and interpretation. *Paediatr Perinat Epidemiol.* 2018, In press.
- Liauw J, Jacobsen G, Larose T, Hutcheon JA. Short interpregnancy interval and poor fetal growth: evaluating the role of pregnancy intention. *Paediatr Perinat Epidemiol.* 2018, In press.
- Hernan MA, Robins JM. Estimating causal effects from epidemiological data. *J Epidemiol Community Health.* 2006;60:578-586.
- Garcia-Albeniz X, Hsu J, Hernan MA. The value of explicitly emulating a target trial when using real world evidence: an application to colorectal cancer screening. *Eur J Epidemiol.* 2017;32:495-500.
- Moons KG, Royston P, Vergouwe Y, Grobbee DE, Altman DG. Prognosis and prognostic research: what, why, and how? *BMJ.* 2009;338:b375.
- Greenland S, Pearl J, Robins JM. Causal diagrams for epidemiologic research. *Epidemiology.* 1999;10:37-48.
- Hernán MA, Hernández-Díaz S, Werler MM, Mitchell AA. Causal knowledge as a prerequisite for confounding evaluation: an application to birth defects epidemiology. *Am J Epidemiol.* 2002;155:176-184.
- Grimshaw J, Campbell M, Eccles M, Steen N. Experimental and quasi-experimental designs for evaluating guideline implementation strategies. *Fam Pract.* 2000;17(Suppl 1):S11-S16.
- Meyer BD. Natural and Quasi-experiments in economics. *J Bus Econ Stat.* 1995;13:151-161.
- Bailey MJ, Hershbein B, Miller AR. The opt-in revolution? Contraception and the gender gap in wages. *Am Econ J Appl Econ.* 2012;4:225-254.
- Yudkin PL, Wood L, Redman CW. Risk of unexplained stillbirth at different gestational ages. *Lancet.* 1987;1:1192-1194.
- Arnold CC, Kramer MS, Hobbs CA, McLean FH, Usher RH. Very low birth weight: a problematic cohort for epidemiologic studies of very small or immature neonates. *Am J Epidemiol.* 1991;134:604-613.
- May S, Bigelow C. Modeling nonlinear dose-response relationships in epidemiologic studies: statistical approaches and practical challenges. *Dose Response.* 2006;3:474-490.
- Royston P, Ambler G, Sauerbrei W. The use of fractional polynomials to model continuous risk variables in epidemiology. *Int J Epidemiol.* 1999;28:964-974.
- Harrell FE Jr, Lee KL, Pollock BG. Regression models in clinical studies: determining relationships between predictors and response. *J Natl Cancer Inst.* 1988;80:1198-1202.
- Dawson NV, Weiss R. Dichotomizing continuous variables in statistical analysis: a practice to avoid. *Med Decis Making.* 2012;32:225-226.
- Conde-Agudelo A, Rosas-Bermudez A, Kafury-Goeta AC. Effects of birth spacing on maternal health: a systematic review. *Am J Obstet Gynecol.* 2007;196:297-308.
- Jones RK, Kost K. Underreporting of induced and spontaneous abortion in the United States: an analysis of the 2002 National Survey of Family Growth. *Stud Fam Plann.* 2007;38:187-197.
- Naimi AI. Studying interpregnancy interval effects using observational data: some cautionary remarks. *BJOG.* 2016;123:1319.
- Conzuelo-Rodríguez G, Naimi AI. The impact of computing interpregnancy intervals without accounting for intervening pregnancy events. *Paediatr Perinat Epidemiol.* 2018;32:141-148.
- Greenland S. Interactions in epidemiology: relevance, identification, and estimation. *Epidemiology.* 2009;20:14-17.
- Kaharuza FM, Sabroe S, Basso O. Choice and chance: determinants of short interpregnancy intervals in Denmark. *Acta Obstet Gynecol Scand.* 2001;80:532-538.
- Stephansson O, Dickman PW, Cnattingius S. The influence of interpregnancy interval on the subsequent risk of stillbirth and early neonatal death. *Obstet Gynecol.* 2003;102:101-108.
- Howards PP, Schisterman EF, Heagerty PJ. Potential confounding by exposure history and prior outcomes: an example from perinatal epidemiology. *Epidemiology.* 2007;18:544-551.
- Weinberg CR. Toward a clearer definition of confounding. *Am J Epidemiol.* 1993;137:1-8.
- Ahrens KA, Thoma M, Copen C, Frederiksen B, Decker E, Moskosky S. Unintended pregnancy and interpregnancy interval by maternal age, National Survey of Family Growth. *Contraception.* 2018;98:52-55.
- Hall JA, Benton L, Copas A, Stephenson J. Pregnancy Intention and Pregnancy Outcome: systematic Review and Meta-Analysis. *Matern Child Health J.* 2017;21:670-704.
- Hernán MA, Robins JM. *Causal Inference*. Boca Raton, FL: Chapman & Hall/CRC, forthcoming; 2018.



37. Schisterman EF, Perkins NJ, Mumford SL, Ahrens KA, Mitchell EM. Collinearity and causal diagrams: a lesson on the importance of model specification. *Epidemiology*. 2017;28:47-53.
38. Ananth CV, Schisterman EF. Confounding, causality, and confusion: the role of intermediate variables in interpreting observational studies in obstetrics. *Am J Obstet Gynecol*. 2017;217:167-175.
39. Forrow L, Taylor WC, Arnold RM. Absolutely relative: how research results are summarized can affect treatment decisions. *Am J Med*. 1992;92:121-124.
40. Naylor CD, Chen E, Strauss B. Measured enthusiasm: does the method of reporting trial results alter perceptions of therapeutic effectiveness? *Ann Intern Med*. 1992;117:916-921.
41. Wesselink AK, Rothman KJ, Hatch EE, Mikkelsen EM, Sorensen HT, Wise LA. Age and fecundability in a North American pre-conception cohort study. *Am J Obstet Gynecol*. 2017;217:667.e661-667.e668.
42. Donovan SJ, Susser E. Commentary: advent of sibling designs. *Int J Epidemiol*. 2011;40:345-349.
43. Richmond RC, Al-Amin A, Smith GD, Relton CL. Approaches for drawing causal inferences from epidemiological birth cohorts: a review. *Early Human Dev*. 2014;90:769-780.
44. Maclure M. The case-crossover design: a method for studying transient effects on the risk of acute events. *Am J Epidemiol*. 1991;133:144-153.
45. 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.
46. Louis GB, Dukic V, Heagerty PJ, et al. Analysis of repeated pregnancy outcomes. *Stat Methods Med Res*. 2006;15:103-126.
47. Zeger SL, Liang KY. An overview of methods for the analysis of longitudinal data. *Stat Med*. 1992;11:1825-1839.
48. Frisell T, Oberg S, Kuja-Halkola R, Sjolander A. Sibling comparison designs: bias from non-shared confounders and measurement error. *Epidemiology*. 2012;23:713-720.
49. Keyes KM, Smith GD, Susser E. On sibling designs. *Epidemiology*. 2013;24:473-474.
50. Sjölander A, Frisell T, Kuja-Halkola R, Öberg S, Zetterqvist J. Carryover effects in sibling comparison designs. *Epidemiology*. 2016;27:852-858.
51. Ahrens KA, Rossen LM, Thoma ME, Warner M, Simon AE. Birth order and injury-related infant mortality in the U.S. *Am J Prev Med*. 2017;53:412-420.
52. Ahrens KA, Thoma ME, Rossen LM, Warner M, Simon AE. Plurality of birth and infant mortality due to external causes in the United States, 2000-2010. *Am J Epidemiol*. 2017;185:335-344.

How to cite this article: Hutcheon JA, Moskosky S, Ananth CV, et al. Good practices for the design, analysis, and interpretation of observational studies on birth spacing and perinatal health outcomes. *Paediatr Perinat Epidemiol*. 2019;33:O15–O24. <https://doi.org/10.1111/ppe.12512>