Promise and Pitfalls of the Sibling Comparison Design in Studies of Optimal Birth Spacing

Jennifer A. Hutcheon and Sam Harper

Correspondence to: Dr. Jennifer Hutcheon, Department of Obstetrics and Gynaecology, Shaughnessy C408A, BC Children's & Women's Hospital, 4500 Oak Street, Vancouver, British Columbia, V6H 3N1, Canada (email: jhutcheon@bcchr.ca)

Corresponding author during editing: Dr. Sam Harper, Department of Epidemiology, Biostatistics, and Occupational Health, Purvis Hall, 1020 Pine Avenue W. Montreal, QC, Canada H3A 1A2 (sam.harper@mcgill.ca) Tel.: 514-398-2856 Fax: 514-398-4503

Author affiliations: Department of Obstetrics and Gynaecology, University of British Columbia, Vancouver, Canada (Jennifer A Hutcheon); Department of Epidemiology, Biostatistics, and Occupational Health, McGill University, Montreal, Canada (Sam Harper)

JAH holds a Canada Research Chair in Perinatal Population Health. SH holds a *Chercheur-Boursier* from the *Fonds de Recherche-Santé Quebec.*

Conflicts of interest: none declared

Running head**:** Sibling Comparison Design in Birth Spacing Studies

Abbreviations: OR, odds ratio

Abstract

Numerous observational studies have shown that infants born after short interpregnancy intervals (the interval between birth and subsequent conception) are more likely to experience adverse perinatal outcomes than infants born following longer intervals. Yet, it remains controversial whether the link between short interpregnancy interval and adverse outcomes is causal, or confounded by factors such as low socio-economic position, inadequate access to healthcare, and unintended pregnancy. Sibling comparison studies, which use a woman as her own control by comparing exposure and outcome status of her different pregnancies (i.e., comparing sibling offspring), have gained popularity as a strategy to reduce confounding by these difficult-to-measure factors that are nevertheless relatively stable within women. A variant of the design, used by Regan *et al.* in this issue (*Am J Epidemiol* 2018;XX:XXX-XXX), is a maternally-matched design based on a single interpregnancy interval per woman. Using real and simulated data, we highlight underappreciated shortcomings of these designs that may limit the validity of study findings. In particular, we illustrate how the single-interval variant appears to derive estimates from comparisons between different mothers, not within mothers. Future studies of optimal birth spacing using sibling comparison designs should examine in detail the potential consequences of these methodological limitations.

Keywords: Birth spacing, interpregnancy interval, confounding, sibling comparison design, preterm birth, adverse perinatal outcome.

Introduction

The World Health Organization recommends that women wait at least 24 months after delivery of a live birth before conceiving another child.[1] These recommendations are informed by a meta-analysis of observational studies showing that short interpregnancy intervals (the interval between delivery and conception of a subsequent child) are associated with increased risks of adverse perinatal outcomes such as preterm birth.[2] Preventing short interpregnancy intervals through improved access to family planning tools could be one strategy to reduce the global burden of perinatal morbidity and mortality.

Yet, it remains controversial whether the observed associations between short interpregnancy interval and adverse perinatal outcomes reflect a causal effect of birth spacing.[3] Short interpregnancy intervals are more common in pregnancies that were unintended, and in women of lower socio-economic position, both of which are also risk factors for adverse perinatal outcomes.[4-6] Given the challenges in measuring and quantifying complex constructs such as pregnancy intention and socio-economic position, especially in cohorts large enough to study rare outcomes such as stillbirth, the confounding in conventional observational studies of interpregnancy interval and adverse perinatal outcomes seems almost intractable.

Promise: The appeal of the sibling comparison design

Sibling comparison studies have emerged as a promising strategy to account for confounding by socio-economic and other risk factors in studies of optimal birth spacing. In this design, a woman is used as her own control by leveraging variation in interpregnancy intervals across her successive pregnancies to estimate the effect on perinatal outcomes. Characteristics that remain constant across a woman's pregnancies, such as maternal ethnicity and some lifestyle factors, cannot confound the exposure-outcome association since there is no variation in the

characteristic within each maternally-matched comparison group. The design is perceived to produce less confounded estimates of the causal effect of short birth spacing on adverse perinatal outcomes, and has been employed by at least five research teams since its first application to birth spacing research in 2014.[7-11] Variants based on comparing cousins, rather than siblings, have also been published.[12]

Pitfalls: Underappreciated limitations of the sibling comparison design

Despite its intuitive appeal, the sibling comparison design introduces its own potential concerns.[13-15] First, the design does not control for confounders that vary across pregnancies, which may be a strong assumption. Although factors such as pregnancy intention and socio-economic position are often considered fixed, this is not the case for all women. For example, Basso *et al.* reported that 28.5% of women in a large population-based study from Denmark changed social status between consecutive births.[16] As a result, control for confounding may not be as complete as believed.

Another concern is that the design improves internal validity at the expense of generalisability. The sibling comparison design necessarily restricts the cohort to women with three or more pregnancies (variation within mothers requires at least two interpregnancy intervals per woman). The extent to which the birth spacing-adverse outcome association among women with three or more pregnancies is comparable to the target population (new mothers planning another pregnancy) is unclear, although one comparison between these two cohorts did not identified major concerns.[7]

Perhaps a more important threat to generalisability comes from the matching in the study design. As with any matched-pairs design, only women with discordant exposure and outcome status will contribute to the overall estimate of effect.[17] However, the magnitude of this reduction in the effective analytic sample does not appear to be well appreciated. For example, in a sibling

comparison study evaluating the Head Start early childhood health and education program, only one fifth of the sibling cohort (1,098/5,355 siblings) was in a family with discordant exposure status (program participation).[18] Moreover, the characteristics of the families included in the effective analytic sample because of their discordant exposure status differed substantially from those excluded because of their children's concordant exposure status.

The reduction in effective analytic sample is even more pronounced in sibling comparison studies of birth spacing, because adverse perinatal outcomes are uncommon. As a result, few sibling pairs will contain a birth with an adverse outcome, meaning that inference in the sibling comparison design is based on a very small fraction of the original cohort. For heterogenous outcomes such as preterm birth,[19] it is possible that the underlying disease etiology differs in women with a preterm birth in only one pregnancy compared with women with recurrent preterm birth (who are excluded). If so, restricting the analytic sample to discordant siblings could alter the estimated effects of preterm birth risk factors from those based on all women with a preterm birth.

Promise: The single-interval sibling comparison design as a novel alternative

In this edition, Regan and colleagues apply a novel variant of the sibling comparison design.[20] In this design, previously used to study the link between interpregnancy interval and autism,[21] only two pregnancies per woman are used to estimate the effect of interpregnancy interval on adverse birth outcomes. Conditional logistic regression is used to model the odds of adverse perinatal outcome in the second pregnancy, with a term for birth order and a product term between birth order and interpregnancy interval. If the product term is positive and significant, this is interpreted to mean that there is an excess risk among births delivered following a short interpregnancy interval (i.e., above and beyond each woman's "baseline" risk of the adverse outcome from her first pregnancy).

The manuscript helps to advance our understanding of the link between short birth spacing and adverse perinatal outcomes in several ways. Its use of this novel design draws attention to generalisability concerns introduced by restricting sibling comparison studies of birth spacing to women with three or more births. Further, the detailed tables and appendices allow a clear examination of how the matched design's restriction to discordant women impacts generalisability. For example, Tables 2 and S3 show the odds of adverse perinatal outcomes using conventional (unmatched) logistic regression. The only difference between the two tables is that Table 2 estimates are based on all multiparous women, whereas Table S3 estimates are restricted to multiparous women with discordant siblings. The adjusted coefficient for interpregnancy interval <6 months associated with preterm birth decreased from 1.67 [95%CI: 1.50, 1.85] in all multiparous women to 1.13 [95%CI: 0.96, 1.32] among women with discordant siblings (i.e., the same cohort used for the matched sibling model). This suggests that much of the difference between the matched and unmatched estimates may reflect not so much an improved control for confounding of the matched design *per se*, but rather, differences in the exposure-outcome relationship between the cohort of all multiparous women and the much smaller cohort of women with discordant outcomes (here, only 9.3% [17,840/192,041] of the original cohort). Providing 'unmatched' estimates for discordant pairs highlights underappreciated concerns regarding generalizability in sibling comparison studies of birth spacing, and serves as a model for researchers reporting results of similar studies.

Pitfalls: Understanding what, exactly, the single-interval sibling comparison design is estimating

Nevertheless, the interpretation of parameters in this single-interval sibling comparison model is not straightforward. Because comparing different interpregnancy intervals within the same woman requires at least 3 pregnancies, restricting to only 2 pregnancies guarantees that

interpregnancy intervals must be identical within women-- a seemingly insurmountable problem in a within-woman design aiming to isolate the effect of inter-pregnancy interval.

To better understand the consequences of having only a single interpregnancy interval per woman, we created a simulated example based on the data reported by the authors for the categories of <6 months and 18-23 months (code to replicate our analysis is in the Web Appendix). We created a population of roughly 20,000 mothers (15% with short interpregnancy intervals), with a 6.2% preterm birth rate. For simplicity, we ignore covariates and only consider binary categories of birth order (second vs. first birth) and interpregnancy interval (IPI; short vs. non-short) as exposures, and a binary outcome (preterm birth). Our **Table 1** shows the distribution of mothers across the resulting 8 possible covariate patterns.

We start by considering the conventional unmatched model of:

(1) $$logit(preterm = 1) = \beta_0 + \beta_1 IPI + \beta_2 birth\ order.$$

We see that short interpregnancy interval is associated with increased odds of preterm birth (odds ratio, OR=exp(0.55)=1.74, see our **Table 2**, Model 1), as well as a negative association between birth order and preterm birth (OR=exp(-0.27)=0.76).

Next, we extend the model to allow for effect measure modification:

(2) $$logit(preterm = 1) = \beta_0 + \beta_1 IPI + \beta_2 birth\ order + \beta_3 IPI * birth\ order$$

Estimates from this model are in our Table 2, Model 2 and demonstrate that the association between short interpregnancy interval and preterm birth varies by birth order, and vice versa. With this model, we can estimate the effect of interpregnancy interval on preterm birth separately for women with second births, as the authors did in their Table 2. The linear combination of coefficients for short interpregnancy interval ($\beta_1 = 0.47$) and the product term ($\beta_3 = 0.18$) generates an odds ratio of OR=exp(0.65)=1.92 for short vs. non-short

interpregnancy interval, which is nearly identical to that the authors report for their unmatched estimate for the <6 months category (OR=1.90).

Now, consider the matched sibling model. The approach used by the authors fits a conditional (matched) model similar to (2) above, except that it only includes terms for birth order (β_2) and the product term (β_3):

$$(3) \qquad logit(preterm = 1 | \, mother = i) = \beta_0 + \beta_2 birth\ order + \beta_3 IPI * birth\ order$$

It omits β_1 because the lack of variation in interpregnancy interval within women means there is no way to estimate this parameter using a model that conditions on mothers. This is immediately evident if one tries to fit a conditional model without the product term (β_3). Because any factors that do not vary within mothers are conditioned out of the likelihood, the parameter for interpregnancy interval is dropped, and we are left with only a conditional estimate of birth order (our Table 2, Model 3).

As previously noted, the effective analytic sample in a matched sibling comparison model comes from the women with discordant outcomes (covariate patterns 2,3,6,7 in our Table 1). If we fit a conditional model to equation (2) in our data, it is apparent that there are not enough degrees of freedom to estimate all three parameters, since the only factor that differs within mothers is birth order. That is, if we restrict to women with a short interpregnancy interval, any terms involving interpregnancy interval (β_1 and β_3) are dropped (our Table 2, Model 4) and we can only obtain an estimate for the effect of increasing birth order (OR=exp(-0.16)=0.86). Similarly, if we try to estimate the effect for women with non-short interpregnancy intervals, the terms involving interpregnancy interval are again dropped and we only get another estimate for birth order (OR=exp(-0.38)=0.68, our Table 2, Model 5).

These models illustrate a core problem with the approach. We want to isolate the unique effect of interpregnancy interval contained in β_3, but to do so, we need to subtract out the effect of birth order, β_2, from model (3). The product term β_3 will only be turned on in women with a short interpregnancy interval (i.e., women for whom IPI is set to 1). However, in these women, we cannot get a clean estimate of the effect of β_2 because the **difference** in the log odds of preterm birth for women in second births ($\beta_0 + \beta_2 + \beta_3$) vs. first births ($\beta_0$) is always $\beta_2 + \beta_3$. The only way to get a clean estimate of β_2 is from women with non-short interpregnancy intervals (for whom the product term β_3 will be turned off, as IPI is set to 0), namely by estimating the difference in the log odds of preterm birth for women with non-short interpregnancy interval in their second birth ($\beta_0 + \beta_2$) vs. first birth (β_0). When we view the model in these terms we can see that the estimate of β_3 the authors report as the "effect" of interpregnancy interval in their matched analysis is simply the difference in the association between birth order and preterm birth among mothers with short interpregnancy interval ($\beta_2 + \beta_3$) vs. mothers with a non-short interpregnancy interval (β_2).

Model 6 in our Table 2 shows the conditional model fit on the whole sample, and it is clear that the estimate of β_3 (OR=exp(0.22)=1.25) that was interpreted as the effect of interpregnancy interval is exactly the difference in the conditional effect of birth order **between** mothers with short (-0.38) vs. non-short IPI (-0.16) who, by definition, must be different mothers. The fact that these estimates come from different mothers means the approach has re-introduced the problem of confounding by between mother factors, and seriously limits any inference about isolating the effect of interpregnancy interval from that of birth order.

Because the conditional model the authors fit does not estimate the effect of interpregnancy interval but rather the difference in the association between birth order and preterm birth for

women with different interpregnancy intervals, this also means that the "null" estimates of interpregnancy interval for SGA and LBW in the authors' Table 3 should not be taken as evidence that there is no interpregnancy interval effect. Rather, it suggests that the associations between birth order and SGA or LBW are just not that different for women with short vs. non-short interpregnancy intervals.

Conclusion

Sibling comparison designs are an attractive alternative to control for difficult-to-measure confounders in studies of short birth spacing and perinatal health. However, methodological issues, such as time-varying confounders and loss of generalisability in the analytic cohort, must be investigated and reported in detail to better understand the extent to which these studies actually provide improved causal estimates. The single-interval design variant used by Regan et al.[20] and others[21], for example, appears to produce estimates generated from between-woman rather than within-woman comparisons, and as such, may not control for confounding by lifestyle or socio-economic factors that remain constant within women as purported. Researchers should continue to explore novel study methodologies (as well as intervention-based approaches such as home visitation programs that encourage longer interpregnancy intervals[22]) to help advance our understanding of the causal effect of short birth spacing on adverse outcomes.

References

1. Department of Reproductive Health and Research. *Report of a WHO Technical Consultation on Birth Spacing: Geneva, Switzerland 13–15 June 2005.* Geneva, Switzerland: World Health Organization; 2006. Retrieved June 20, 2018 http://apps.who.int/iris/bitstream/10665/69855/1/WHO_RHR_07.1_eng.pdf.
2. Conde-Agudelo A, Rosas-Bermudez A, Kafury-Goeta AC. Birth spacing and risk of adverse perinatal outcomes: a meta-analysis. *JAMA.* 2006;295(15):1809-1823.
3. Klebanoff MA. Interpregnancy Interval and Pregnancy Outcomes: Causal or Not? *Obstet Gynecol.* 2017;129(3):405-407.
4. Ahrens KA, Thoma M, Copen C, et al. Unintended pregnancy and interpregnancy interval by maternal age, National Survey of Family Growth. [published online ahead of print March 1, 2018] *Contraception.* (doi. 10.1016/j.contraception.2018.02.013).
5. Hall JA, Benton L, Copas A, et al. Pregnancy Intention and Pregnancy Outcome: Systematic Review and Meta-Analysis. *Mat Child Health J.* 2017;21(3):670-704.
6. Kaharuza FM, Sabroe S, Basso O. Choice and chance: determinants of short interpregnancy intervals in Denmark. *Acta Obstet Gynecol Scand.* 2001;80(6):532-538.
7. Hanley GE, Hutcheon JA, Kinniburgh BA, et al. Interpregnancy Interval and Adverse Pregnancy Outcomes: An Analysis of Successive Pregnancies. *Obstet Gynecol.* 2017;129(3):408-415.
8. Ball SJ, Pereira G, Jacoby P, et al. Re-evaluation of link between interpregnancy interval and adverse birth outcomes: retrospective cohort study matching two intervals per mother. *BMJ.* 2014;349:g4333.
9. Shachar BZ, Mayo JA, Lyell DJ, et al. Interpregnancy interval after live birth or pregnancy termination and estimated risk of preterm birth: a retrospective cohort study. *BJOG.* 2016;123(12):2009-2017.

10. Koullali B, Kamphuis EI, Hof MH, et al. The Effect of Interpregnancy Interval on the Recurrence Rate of Spontaneous Preterm Birth: A Retrospective Cohort Study. *Am J Perinatol.* 2017;34(2):174-182.

11. Perin J, Walker N. Potential confounding in the association between short birth intervals and increased neonatal, infant, and child mortality. *Glob Health Action.* 2015;8:29724.

12. Class QA, Rickert ME, Oberg AS, et al. Within-Family Analysis of Interpregnancy Interval and Adverse Birth Outcomes. *Obstet Gynecol.* 2017;130(6):1304-1311.

13. Frisell T, Oberg S, Kuja-Halkola R, et al. Sibling comparison designs: bias from non-shared confounders and measurement error. *Epidemiol.* 2012;23(5):713-720.

14. Sjolander A, Frisell T, Kuja-Halkola R, et al. Carryover Effects in Sibling Comparison Designs. *Epidemiol.* 2016;27(6):852-858.

15. Keyes KM, Smith GD, Susser E. On sibling designs. *Epidemiol.* 2013;24(3):473-474.

16. Basso O, Olsen J, Knudsen LB, et al. Low birth weight and preterm birth after short interpregnancy intervals. *Am J Obstet Gynecol.* 1998;178(2):259-263.

17. Maclure M. The case-crossover design: a method for studying transient effects on the risk of acute events. *Am J Epidemiol.* 1991;133(2):144-153.

18. Grosz MZ, Miller DL, Shenhav N. Long Term Effects of Head Start: New Evidence from the PSID. *Working paper, University of California Davis Department of Economics.* 2016. Retrieved June 25, 2018 http://www.sole-jole.org/17283.pdf

19. Savitz DA, Blackmore CA, Thorp JM. Epidemiologic characteristics of preterm delivery: etiologic heterogeneity. *Am J Obstet Gynecol.* 1991;164(2):467-471.

20. Regan AK, Ball SJ, Warren JL, et al. A population-based matched sibling analysis estimating the association between first interpregnancy interval and birth outcomes. Am J Epidemiol. 2018;XXX(XX):XXX-XXX.

21. Cheslack-Postava K, Liu K, Bearman PS. Closely spaced pregnancies are associated with increased odds of autism in California sibling births. *Pediatrics.* Feb 2011;127(2):246-253.

22. Goyal NK, Folger AT, Hall ES, Greenberg JM, Van Ginkel JB, Ammerman RT. Home visiting for first-time mothers and subsequent pregnancy spacing. *J Perinatol.* 2017;37(2):144-149.

Table 1. Example cohort of single-interval sibling comparison design

Mother covariate pattern	Birth order	Interpregnancy interval (months)	Preterm birth	Discordant outcome	N
1	1st	<6	No	No	2,435
	2nd	<6	No	No	2,435
2	**1st**	**<6**	**No**	**Yes**	**184**
	2nd	**<6**	**Yes**	**Yes**	**184**
3	**1st**	**<6**	**Yes**	**Yes**	**215**
	2nd	**<6**	**No**	**Yes**	**215**
4	1st	<6	Yes	No	74
	2nd	<6	Yes	No	74
5	1st	18–23	No	No	15,496
	2nd	18–23	No	No	15,496
6	**1st**	**18–23**	**No**	**Yes**	**598**
	2nd	**18–23**	**Yes**	**Yes**	**598**
7	**1st**	**18–23**	**Yes**	**Yes**	**874**
	2nd	**18-23**	**No**	**Yes**	**874**
8	1st	18-23	Yes	No	234
	2nd	18-23	Yes	No	234

Table 2. Regression model coefficients from matched and unmatched models of interpregnancy interval and preterm birth.

Model	Observations	Parameter							
		Intercept		IPI		Birth order		IPI x Birth order	
		β_0	95% CI	β_1	95% CI	β_2	95% CI	β_3	95% CI
1 (unmatched)	40220	−2.69	−2.75, 2.63	0.55	0.45, 0.65	−0.27	−0.35, −0.18		
2 (unmatched)	40220	−2.68	−2.74, −2.62	0.47	0.34, 0.61	−0.30	−0.40, −0.21	0.18	−0.02, 0.38
3 (matched)	3742	n/a		0.00	[dropped]	−0.33	−0.42, −0.24		
4 (matched, short IPI mothers only)	2944	n/a		0.00	[dropped]	−0.16	−0.35, 0.04	0.00	[dropped]
5 (matched, non-short IPI mothers only)	798	n/a		0.00	[dropped]	−0.38	−0.48, −0.28	0.00	[dropped]
6 (matched, full sample)	3742	n/a		0.00	[dropped]	−0.38	−0.48, −0.28	0.22	0.00, 0.45

Interpregnancy interval (IPI)