Human Reproduction Vol.22, No.8 pp. 2080–2083, 2007 doi:10.1093/humrep/dem151
Advance Access publication on June 22, 2007

DEBATE—CONTINUED

Should we adjust for gestational age when analysing birth weights? The use of z-scores revisited

Ilse Delbaere[1,4], Stijn Vansteelandt[2], Dirk De Bacquer[3], Hans Verstraelen[1], Jan Gerris[1], Petra De Sutter[1] and Marleen Temmerman[1]

[1]Department of Obstetrics and Gynaecology, Ghent University, De Pintelaan 185, B-9000 Ghent, Belgium; [2]Department of Applied Mathematics and Informatics, Ghent University, Krijgslaan 281 S9, B-9000 Ghent, Belgium; [3]Department of Public Health, Ghent University, De Pintelaan 185, B-9000 Ghent, Belgium

[4]Correspondence address. Tel: +32 (0)9 240 48 53; Fax: +32 (0)9 240 38 31; E-mail: ilse.delbaere@ugent.be

Birth weight is the single most important risk indicator for neonatal and infant mortality and morbidity, which has led to the idiom that 'every ounce counts'. Birth weight in turn, however, tends to vary widely across populations as a result of differential fetal growth velocity with such demographic factors as ethnicity, maternal and paternal height and altitude of residence. Accordingly, it has been acknowledged that the appraisal of birth weight should rely on its position relative to the birth weight distribution of the background population. This is commonly done by standardizing birth weight through its deviation from the population mean in the given gestational age stratum, as can be obtained from population-customized birth weight nomograms. This issue was recently revisited in '*Human Reproduction*' through a plea for reporting birth weight as *z*-scores. In this article, we argue that adjustment for factors, such as gestational age, which may lie on the causal pathway from exposures present at the time of conception [e.g. single-embryo transfer (SET) versus double-embryo transfer (DET)] to birth weight, may induce bias, regardless of whether the adjustment happens via stratification, regression or through the use of *z*-scores.

Keywords: birth weight; gestational age; z-score

Should we adjust for gestational age when analysing birth weights?

To support the contention that adjustment for gestational age in analysing birth weight differences between two populations may cause bias, we will apply so-called causal diagrams or 'causal directed acyclic graphs'. The theory behind these diagrams has been described extensively by Robins, Hernan and others (Robins, 2001; Hernan *et al*., 2004).

Briefly, causal diagrams display exposures, outcomes, confounders and intermediate variables schematically by means of arrows that represent the possibility of a direct causal effect of the variable at the tail of the arrow on the variable at the head of the arrow (see Fig. 1). For the arrow from an exposure to an outcome in the diagram to carry a causal meaning, such diagram must additionally include all common causes of both.

To understand the impact of adjustment for gestational age in our previous paper (De Sutter *et al*., 2006), we used such causal diagrams (see Fig. 2). In this study, we compared outcomes (gestational age at birth and birth weight) between children after SET and singletons after DET. There were no intermediate factors hindering the relationship between SET/DET and the outcome, but we adjusted for the following confounders: parity, maternal age, indication for assisted reproduction, applied assisted reproductive technique (ART), time of embryo transfer, cycle rank, embryo quality and compulsion to apply SET.

Although all analyses were adjusted for these confounders, we will remove them for simplicity from all further diagrams in order to clarify why adjustment for gestational age is not warranted when estimating the effect of SET/DET on birth weight. The simplest possible causal diagram which then results, is presented in Fig. 3. In this diagram, it is assumed that SET/DET may indirectly affect birth weight by altering gestational length. Under this assumption, birth weight is not associated with number of embryos transferred (SET versus DET) among babies with the same gestational length, which contradicts our (empirical) findings. Specifically, we estimated an odds ratio (OR) of 2.99 (95% confidence interval (CI) 1.06–2.94) for low birth weight in DET singletons when compared to SET singletons; adjustment for gestational age resulted in an OR of 2.44 (95% CI 1.27–4.67).

These findings suggest the presence of a direct effect of SET/DET on birth weight, as depicted in Fig. 4, with an additional indirect effect through gestational age.

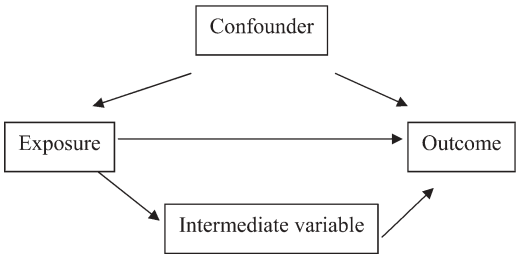

Figure 1: Directed acyclic graph

However, it is likely that besides the number of embryos transferred (SET versus DET) other risk factors also jointly impinge on birth weight and gestational age (e.g. maternal age). In the presence of such risk factors, adjustment for gestational age may induce spurious associations between SET/DET and birth weight, regardless of the presence of a direct effect of SET/DET on birth weight (Robins, 2001; Hernan *et al*., 2004).

Figure 5 depicts a situation where (as is likely the case in practice) there are possibly unmeasured common risk factors for gestational age and birth weight, e.g. genetic predetermination. Under this setting, the data cannot distinguish between a direct effect of SET/DET on birth weight and the presence of unmeasured risk factors jointly affecting gestational age and birth weight (Robins, 2001; Hernan *et al*., 2004). Given the bias that may result from inappropriate adjustment, we deliberately refrained from adjusting for gestational age. As such, under the realistic assumptions of Fig. 5, we have estimated the overall effect of SET/DET on birth weight (i.e. the combination of a direct effect and an indirect effect along gestational age) to be 3.38 (adjusted OR, 95% CI 1.86–6.12).

A similar problem of adjustment for a variable, i.e. both affected by the exposure and shares common causes with the outcome previously gave rise to the birth weight paradox, which has been a topic of debate for several decades (Hernandez-Diaz *et al*., 2006). Specifically, many studies have reported a beneficial effect of smoking on perinatal mortality in low-birth babies (i.e. after adjusting for birth weight). Hernandez-Diaz *et al*. concluded that this was due to inappropriate adjustment for birth weight and indeed observed a harmful effect of smoking on perinatal mortality without such adjustment.

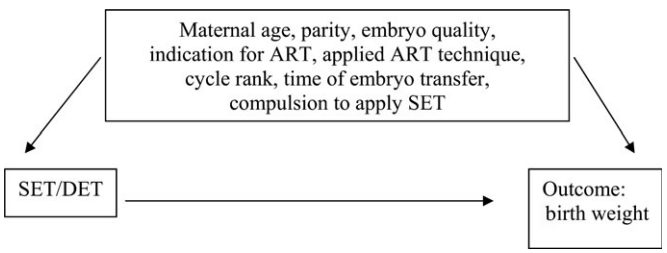

Figure 2: Directed acyclic graph for the analyses assessing outcome differences between singletons after SET and singletons after DET

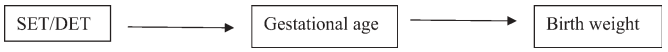

Figure 3: Directed acyclic graph, representing gestational age as only confounder in the pathway between SET versus DET and birth weight

The use of *z*-scores: a remedy?

In the previous discussion, we have shown that adjustment for gestational age is unwarranted when inferring the effect of an exposure A, which arises at conception (e.g. $A = 0$ in SET, $A = 1$ in DET), on birth weight Y. Assuming for simplicity of exposition that both exposed subgroups are comparable at conception (as would be the case if the number of transferred embryos were randomized), then the term β_1 in the linear regression model:

$$E(Y|A) = \beta_0 + \beta_1 A \tag{1}$$

expresses the effect of exposure A on average birth weight. Here, E(Y|A) denotes average birth weight among children with the given exposure A.

Land (2006) suggests replacing birth weight by a z-score, standardizing for gestational age. Specifically, she proposes, replacing Y with the z-score $Y^* = Y - E^*(Y|X)$, where $E^*(Y|X)$ denotes the average/median birth weight among children with gestational age X in the reference population and where, for academic purposes and without loss of generality, we assume a constant standard deviation (and therefore ignore it) in birth weight when calculating z-scores. We will now examine whether such use of z-scores is valid for estimating the effect of exposure A on birth weight Y, i.e. we will investigate whether the resulting analysis yields the same effect size β_1.

The suggested use of z-scores is tantamount to fitting a regression model of the form.

$$E(Y^*|A) = \beta_0^* + \beta_1^* A \tag{2}$$

This use of z-scores is not valid because β_1^* in model (2) does not represent the causal effect of exposure A on average birth weight. This is because β_1^* may differ from zero, even when there is no causal effect of exposure A on birth weight Y (i.e. even when β_1 equals zero). Indeed, the standardization term $E^*(Y|X)$ in the z-score varies with gestational age (X), which may itself be associated with exposure A, even when birth weight is not associated with exposure A.

To gain insight, we will support our claim with a numerical example (see Table 1). Using the data of De Sutter *et al*. (2006), we estimate an average birth weight of 3324.6 g in the SET population, from which we conclude that β_0 in model (1) equals $E(Y|A = 0) = 3324.6$. Likewise, we estimate an average birth weight of 3204.3 g in the DET population, from which

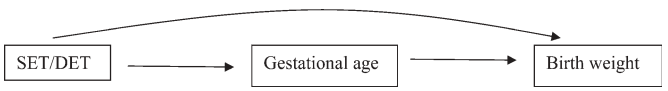

Figure 4: Directed acyclic graph, representing gestational age as a mediator in the pathway between SET versus DET and birth weight

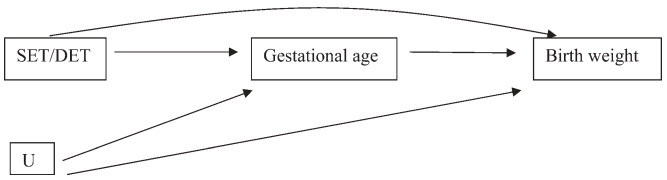

Figure 5: Directed acyclic graph, representing gestational age as a mediator in the pathway between SET versus DET and birth weight and unknown factors having an effect on both gestational age and birth weight

we conclude that the change in average birth weight is $\beta_1 = E(Y|A=1) - E(Y|A=0) = 120.2$ g. This reflects the effect of SET/DET on average birth weight.

For numerical simplicity, let us dichotomize gestational age into preterm born babies ($X=0$) and term born babies ($X=1$). Assume that the average weight of preterm born children after SET equals 3010 and 2687.8 g in children after DET. Then the average birth weight of preterm born children, regardless of the number of embryo's transferred, equals $E(Y|X=0) = 2848.9$ g. Assume furthermore that the average birth weight of term born children after SET equals 3639.0 and 3720.8 g after DET. Then the average birth weight of term born children equals $E(Y|X=1) = 3679.9$ g. Finally, the proportion preterm born children is 6% after SET and 10% after DET (De Sutter *et al*., 2006). The above data now allow us to calculate what effect size we would find through the use of *z*-scores [i.e. by fitting model (2)] and to assess whether this deviates from the true effect size of 120.2 g. Specifically, standard probability calculus shows that.

$$\begin{aligned} E(Y^*|A=0) &= E(Y|A=0) - P(X=1|A=0)E(Y|X=1) \\ &\quad - P(X=0|A=0)E(Y|X=0) \\ &= 3324.6 - 0.94^*3679.9 - 0.06^*2848.9 \\ &= -305.4 = \beta_0^* \end{aligned}$$

where $P(X{=}1|A=0)$ is the proportion of term children after SET.

Likewise,

$$\begin{aligned} E(Y^*|A=1) &= 3204.3 - 0.90^*3679.9 - 0.10^*2848.9 \\ &= -392.5 \end{aligned}$$

It follows that the difference between *z*-scores for birth weight, i.e. β_1^* in model (2), equals $-305.4 + 392.5 = 87.1$, which differs substantially from the real effect of 120.2 g that we initially found.

It follows from the above discussion that a comparison of *z*-scores for pregnancy outcomes between women who were differently exposed at conception, does not yield valid inferences for the effect of exposure on outcome when the *z*-score standardizes outcome w.r.t. factors, such as gestational age, that may themselves be influenced by the exposure. In the Appendix, we show similarly that such comparisons of *z*-scores also fail to remove bias due to confounding factors that were measured at the time of conception, even when the *z*-score standardizes outcome according to those factors. Adjusting comparisons of pregnancy outcomes for the *z*-score for gestational age has no effect and thus yields valid inferences (see the Appendix). However, this is only true when the adjustment happens through linear regression models and when the standardization in the *z*-score is based on subtracting means (and not medians) and happens with relation to the study population (and not some other reference population).

Table 1: Mean birth weight in the SET versus DET population

	SET population	DET population
Average birth weight (g)	3324.6	3204.3
Birth weight in preterm born babies (g)	3010.0	2687.8
Birth weight in term born babies (g)	3639.0	3720.8

Conclusion

In summary, stratifying on an intermediate variable i.e. both affected by the exposure and shares common causes with the outcome is invalid as this may induce bias. Specifically, adjustment for gestational age is generally inappropriate when assessing the effect of reproductive technologies on birth weight outcomes, regardless of whether the adjustment happens through regression models or through the use of *z*-scores. In that case (i.e. without adjustment for such intermediate variable), the estimated effect refers to the overall effect of reproductive technologies on birth weight outcomes, which may act directly, or indirectly by modifying gestational age. The use of *z*-scores can be useful, e.g. in studies of malnutrition in African children during their first year of life. In that case, comparison of *z*-scores of body weight may be useful when the standardization happens with relation to Western European children of the same gender and age. Land (2006) suggested the use of *z*-scores in order to report on birth weight in a standardized way. Since the above arguments refute this recommendation, we suggest reporting differences in incidences of low-birth weight in studies comparing outcome differences between different populations. Comparison of incidences is interesting in terms of clinical relevance; however, since binary variables often imply information loss, it is recommended to provide means of, e.g. birth weight as well. We conclude by advising for the use of causal diagrams in empirical research and against adjustment for factors that lie on the causal pathway from exposures occurring at conception on pregnancy outcome, unless those factors share no common causes with pregnancy outcome.

References

Cole SR, Hernan MA. Fallibility in estimating direct effects. *Int J Epidemiol* 2002;**31**:163–165.

De Sutter P, Delbaere I, Gerris J, Verstraelen H, Goetgeluk S, Van der Elst J, Temmerman M, Dhont M. Birth weight of singletons after assisted reproduction is higher after single than after double embryo transfer. *Hum Reprod* 2006;**21**:2633–2637.

Hernan MA, Hernandez-Diaz S, Robins JM. A structural approach to selection bias. *Epidemiology* 2004;**15**:615–625.

Hernandez-Diaz S, Schisterman EF, Hernan MA. The birth weight paradox uncovered? *Am J Epidemiol* 2006;**164**:1115–1120.

Land JA. How should we report on perinatal outcome? *Hum Reprod* 2006;**21**:2638–2639.

Robins JM. Data, design, and background knowledge in etiologic inference. *Epidemiology* 2001;**12**:313–320.

Submitted on January 23, 2007; resubmitted on April 16, 2007; accepted on April 20, 2007