
Paediatric and Perinatal Epidemiology

294 doi: 10.1111/ppe.12284

Methodology article

Maternal Weight Gain During Pregnancy: Comparing Methods to Address Bias Due to Length of Gestation in Epidemiological Studies

Stefanie N. Hinkle,[a] Emily M. Mitchell,[a] Katherine L. Grantz,[a] Aijun Ye,[b] Enrique F. Schisterman[a]

[a]*Epidemiology Branch, Division of Intramural Population Health Research, Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, MD*

[b]*Glotech, Inc, Rockville, MD*

Abstract

Background: Studies examining total gestational weight gain (GWG) and outcomes associated with gestational age (GA) are potentially biased. The *z*-score has been proposed to mitigate this bias. We evaluated a regression-based adjustment for GA to remove the correlation between GWG and GA, and compared it to published weight-gain-for-gestational-age *z*-scores when applied to a study sample with different underlying population characteristics.

Methods: Using 65 643 singleton deliveries to normal weight women at 12 US clinical sites, we simulated a null association between GWG and neonatal mortality. Logistic regression was used to estimate approximate relative risks (RR) of neonatal mortality associated with GWG, unadjusted and adjusted for GA, and the *z*-score, overall and within study sites. Average RRs across 5000 replicates were calculated with 95% coverage probability to indicate model bias and precision, where 95% is nominal.

Results: Under a simulated null association, total GWG resulted in a biased mortality estimate (RR = 0.87; coverage = 0%); estimates adjusted for GA were unbiased (RR = 1.00; coverage = 94%). Quintile-specific RRs ranged from 0.97–1.03. Similar results were observed for site-specific analyses. The overall *z*-score RR was 0.97 (84% coverage) with quintile-specific RRs ranging from 0.64–0.90. Estimates were close to 1.0 at most sites, with coverage from 70–94%. Sites 1 and 6 were biased with RRs of 0.66 and 1.43, respectively, and coverage of 70% and 80%.

Conclusions: Adjusting for GA achieves unbiased estimates of the association between total GWG and neonatal mortality, providing an accessible alternative to the weight-gain-for-gestational-age *z*-scores without requiring assumptions concerning underlying population characteristics.

Keywords: *Bias, Epidemiology, Weight gain, Pregnancy, Gestational age.*

Gestational weight gain is as an important predictor of maternal and child outcomes.[1] Total gestational weight gain, the difference in weight at delivery and prepregnancy weight, is a summary measure of weight accumulation throughout pregnancy and is inherently associated with gestational age. As demonstrated by Hutcheon *et al*.[2], studies examining the association between total weight gain and preterm delivery, or other outcomes associated with gestational age (e.g. neonatal mortality, birthweight, child development) are potentially biased when they fail to appropriately account for gestational age.

Recently, gestational weight gain *z*-scores have been proposed as a means to address the correlation between total weight gain and gestational age at delivery.[3,4] The published *z*-score charts provide gestational weight gain means and standard deviations as well as selected percentiles by gestational age. These charts were created using a population with repeated measures of weight gain throughout pregnancy with the goal of applying them to a sample in

Correspondence:
Enrique F. Schisterman, Division of Intramural Population Health Research, *Eunice Kennedy Shriver* National Institute of Child Health and Human Development, National Institutes of Health, 6100 Executive Blvd Room 7B03, Bethesda, MD 20892, USA.
E-mail: schistee@mail.nih.gov

which only total weight gain is available to obtain a gestational age standardised exposure for epidemiological analyses. While *z*-scores can help mitigate the correlation between two variables when they adequately represent the underlying population they valid estimation of the gestational-age-specific weight gain means and variances for a *z*-score chart requires additional assumptions. These assumptions include a validation sample representative of the study population and correct specification of the model underlying the relationship between gestational age and weight gain, including the mean, variance, link function, and functional form of the covariates and any potential confounders. The implications of violating these assumptions, such as applying *z*-scores derived from one population to a fundamentally different one, have not been fully evaluated.

An alternative method for addressing correlation between total weight gain and gestational age is regression-based adjustment for gestational age. However, this straightforward approach has also not been evaluated in the context of gestational weight gain and in the past its validity has created controversy.[5–7] There was concern that adjusting for gestational age may induce collider stratification bias by adjusting for a potential intermediate between weight gain and mortality.[7] The adjustment approach has the advantage that researchers can specify the model to fit the structure of the gestational weight gain and gestational age relationship in their observed data.

In this paper, we use directed acyclic graphs (DAGs) and simulation approaches to evaluate the confounding due to gestational age in studies of assessing the relation between gestational weight gain and pregnancy outcomes associated with gestational age. First, we use DAGs to describe the correlation between weight gain and gestational age longitudinally across gestation. This can help clarify assumptions about the directionality of the relations between variables at the cross-sectional time point of delivery, the most common form of perinatal variables. Second, we utilise analytical and simulation approaches to assess the implications and potential biases from using simplified composite measures of total weight gain and gestational age at delivery as related to neonatal mortality, an outcome where critical data gaps remain in the literature.[1] We compare the approach of adjusting for gestational age in a model of total gestational weight to applying previously published *z*-score reference chart values for total gestational weight gain.[3] By using a large, diverse cohort, comprising several sites from across the United States, we evaluate the impact of mis-specifying the distributions of weight gain and gestational age, a key assumption of the *z*-score approach.

Directed acyclic graphs

Gestational weight gain is a time-dependent variable that is often treated as a fixed variable of total weight gain at delivery, as data on total weight gain are more often collected than repeated measures of maternal weight gain. In reality, total gestational weight gain represents the summation of maternal weight gain at each week of gestation culminating in a measure of cumulative weight gain (or loss). Similarly, gestational age at delivery is the summation of indicator variables denoting whether or not the baby is still *in utero* at each gestational week (t). In addition, birthweight is the summation of foetal weight accumulated longitudinally across gestation. The DAG in Figure 1 displays a simplification of these longitudinal time-dependent variables at intermittent gestational ages. We assume that there are no unmeasured confounders and that the DAG is complete. Whether or not a woman is still pregnant (i.e. the baby is *in utero*) impacts her subsequent maternal and foetal weight, essentially through having a longer time to accumulate weight. Also, maternal weight impacts subsequent foetal growth as well as whether or not the woman may deliver at any particular time point. Lastly, foetal weight impacts subsequent maternal weight gain and whether or not a woman will deliver, potentially acting through obstetrical interventions, (e.g. early delivery for a growth restricted foetus to prevent stillbirth), but this is beyond the scope of this diagram. The right side of the diagram shows that gestational age at delivery, total gestational weight gain and birthweight are the summary measures of the longitudinal process "*in utero*" time (IU_t), maternal weight (MW_t) and foetal weight (FW_t), respectively. Lastly, we assume that IU_t and MW_t directly impact neonatal mortality.

Most epidemiological studies only have data on total gestational weight gain at delivery. Figure 2a displays the simplified version of Figure 1 by including only summary measures for the longitudinal processes. In this DAG, a critical point is that when simplified to the cross-section at delivery there is no direct relation between total weight gain and

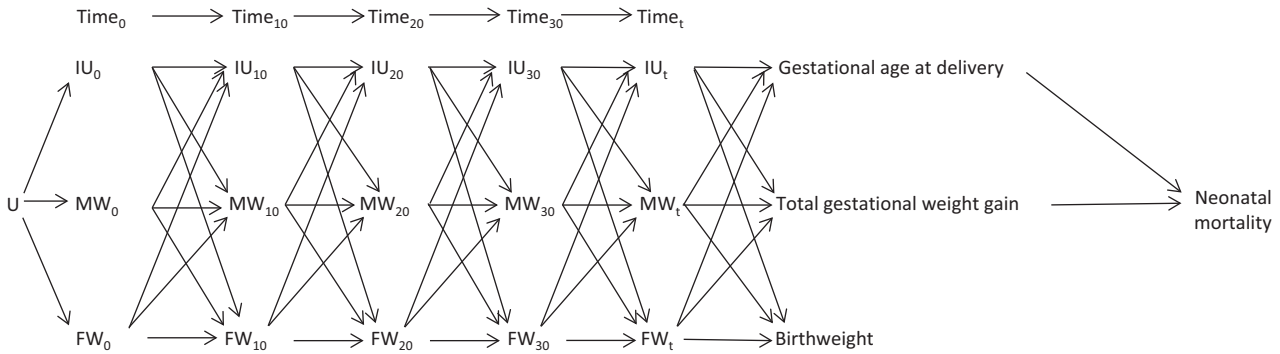

Figure 1. Directed acyclic graph representing the longitudinal relations between maternal weight gain, foetal weight, and gestational age. Where IU represents if the baby is still *in utero*, MW represents maternal weight gain including maternal, foetal and placental tissue and fluid expansion, and FW represents foetal weight.

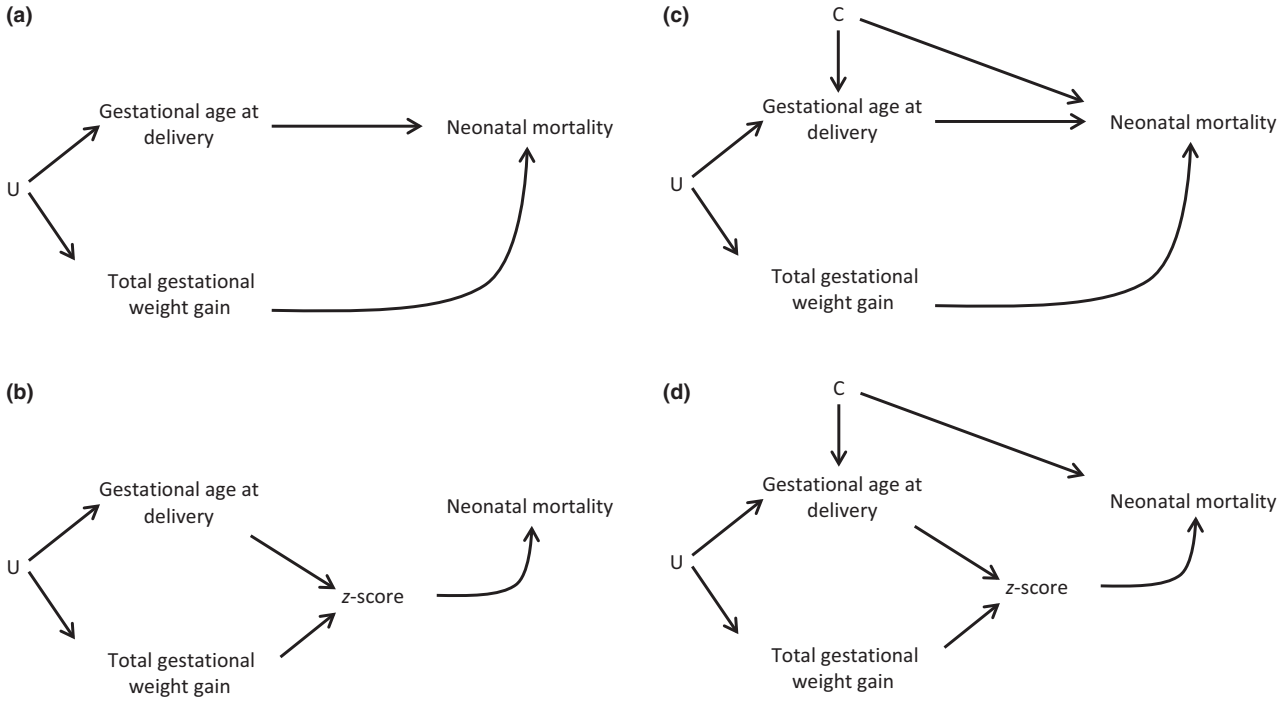

Figure 2. Directed acyclic graphs representing the association between total gestational weight gain (panel a, c) or the weight-gain-for-gestational-age *z*-score (panel b, d) and neonatal mortality. Where "U" represents the longitudinal feedback trajectory of maternal weight gain and gestational age across gestation and "C" represents a confounder of the gestational age at delivery and neonatal mortality relation.

gestational age at delivery, but rather a "U" representing the unobserved longitudinal processes and feedbacks between weight gain and gestational age. This is in contrast to previous assumptions that the summary measure of total gestational weight gain is directly impacted by the summary measure of gestational age at delivery or vice versa, which fails to account for the past time-dependent confounding affected by prior exposures. In this DAG, adjusting for gestational age at delivery blocks the backdoor path through U, removing confounding.

Figure 2b displays a DAG representing the assumptions made by the weight-gain-for-gestational-age *z*-score. As gestational age at delivery was used in the transformation to create the *z*-score there is an arrow between both gestational age and total weight gain and the *z*-score. Taking the approach of modelling the *z*-score as the exposure of interest implies that the effects of gestational weight gain and gestational age on neonatal mortality are assumed to be entirely contained in the *z*-score. The *z*-score is shown as the exposure of

interest to represent its suggested use in epidemiologic studies.[3]

In Figures 2c–d, we build on the DAGs described above to show an additional scenario of interest which includes a confounder, C, of the gestational age at delivery and neonatal mortality relation. Under the scenario that C is measured (e.g. maternal age), an unbiased estimate of the total weight gain and neonatal mortality relation can be achieved by adjusting for C. However, if C is unmeasured (e.g. genetics), under the scenario in Figure 2c, collider bias may be induced when conditioning on gestational age at delivery,[8] while under the *z*-score model in Figure 2d, C remains a confounder of the *z*-score neonatal mortality relation. Both scenarios displayed in Figures 2c–d result in open paths that can lead to biased estimates when C is unmeasured, where the magnitude of the bias depends on the strength of the relationship between C, gestational age at delivery and neonatal mortality, as shown algebraically in Appendix A. It is important to note that collider bias generally tends to be smaller in magnitude than confounding bias,[9,10] but this tendency has yet to be evaluated under these causal structures, which is beyond the scope of this paper.

Methods

We utilised data from the Consortium on Safe Labor (CSL) to compare effect estimates from models that use total gestational weight gain with adjustment for gestational age, to models that apply the weight-gain-for-gestational-age *z*-score to assess the total effect of total weight gain on neonatal mortality. The CSL was comprised of 12 US hospital-based sites (2002–2008).[11] Data were extracted from maternal and neonatal electronic medical records. All study procedures were reviewed and approved by each participating site's Institutional Review Board.

For this analysis, we utilised data from the first singleton birth with information available on prepregnancy weight, height, gestational age, birthweight, and neonatal mortality (n = 121 922). We limited our analysis to normal weight mothers (prepregnancy body mass index (kg/m^2) of 18.5–24.9; (53.9%), to avoid the potential interaction with prepregnancy weight status, and to mothers who delivered between 24 and 40 weeks (n = 65 669). We limited the data to deliveries ≤40 weeks as the weight-gain-for-gestational-age *z*-score chart stops at 40 weeks.[3]

Total gestational weight gain was calculated as the difference between a mother's prepregnancy weight as recorded on her medical chart and her weight recorded at her time of delivery. A corresponding weight-gain-for-gestational-age *z*-score was calculated for each woman based on published *z*-score charts.[3]

As a weight gain below −9.18 kg does not produce a valid *z*-score, we excluded potentially implausible weight gain observations <−9.18 (n = 8; 0.01%) or >50 kg (n = 18; 0.03%) leaving a total of 65 643 deliveries. We categorised total weight gain and *z*-scores into quintiles. Gestational age at delivery was based on the best obstetric estimate available. Neonatal mortality was obtained from the neonatal hospital record.

Our goal was to assess the different methods under a null scenario in which the true association between weight gain and neonatal mortality was known (RR = 1.0) and thus departures from an estimated RR = 1.0 represent bias. To achieve this, we removed the observed neonatal mortality variable and generated a new outcome for each maternal-neonatal pair from a Bernoulli distribution with probability commensurate with the gestational-age specific rates of neonatal mortality observed in the entire CSL dataset. Based on the original CSL distribution, the probability of mortality at 24 weeks was 20% and decreased to 0.04% among neonates born at 40 weeks. Thus in the simulated data, the original correlation between gestational age and gestational weight gain was retained, but the direct link between gestational weight gain and mortality was broken and neonatal mortality was only directly associated with gestational age at delivery. We repeated the process 5000 times, each time randomly generating a new neonatal mortality for each observation. In supplemental simulations (data not shown), we generated the simulated outcome based on site-specific neonatal mortality rates by gestational age to encompass potential site-specific effects on mortality rates. Results from this scenario were consistent with the results presented below.

Relative risks of neonatal mortality associated with gestational weight gain were approximated by calculating odds ratio estimates under a logistic regression model.[12] The first set of models used total gestational weight gain as the exposure of interest, unadjusted and adjusted for gestational age at delivery. The second set of models used the weight-gain-for-gestational-age *z*-score as the exposure of interest. Total weight gain and the *z*-score were examined as both a linear function and as quintiles. We estimated

the linear models by study site because the underlying distribution of weight gain and gestational age may vary across site and this allowed us to assess the *z*-score related assumption of correct specification of the underlying the relationship between gestational age and weight gain. The average risk across the 5000 replicates was calculated based on $\exp(\hat{\beta})$. The coverage probability was calculated as an indicator of the model's ability to provide unbiased estimates of the risk and corresponding estimated standard error, where nominal coverage is 95%. Departures from 95% are indicative of bias in the point estimate or its standard error. Empirical standard error of $\hat{\beta}$ was calculated as an indicator of model precision.

Results

As expected in the United States, maternal demographics of the women in this cohort varied greatly by study site (Table 1). For example, site 6 was predominantly non-Hispanic white (81.5%), while site 12 was predominantly Hispanic (58.0%). The proportion of married women varied across sites as did the per cent of women who smoked prior to pregnancy. On average, the normal weight women in the cohort gained 15.19 kg (standard deviation (5.70)) across pregnancy. There was approximately a 2 kg discrepancy in total weight gain between the sites with the lowest (14.3 kg) and highest (16.2 kg) average total weight gain. Total weight gain was positively correlated with gestational age at delivery at all sites and correlations ranged from 0.07 at site 6 to 0.24 at site 1 ($P < 0.001$) (data not shown). The average weight-gain-for-gestational-age *z*-score exceeds the standard normal implied by the *z*-statistic with an overall mean of −0.18 and a standard deviation of 1.07. The mean ranged from −0.01 to −0.32 depending on study site, indicating that on average the women in the cohort gained less than the women from which the *z*-score charts were developed. Across the 5000 simulations the average neonatal mortality rate was 0.22% (0.04) and was stable by site.

Table 2 displays the simulation study results for the full cohort examining total gestational weight gain unadjusted and adjusted for gestational age at delivery and the *z*-score. The simulated true effect was no association between weight gain and neonatal mortality. As expected, when total gestational weight gain was the exposure of interest, the overall results were biased away from the null (average RR = 0.87) with 0% coverage, indicating that over the 5000 replicates of the simulations the estimated 95% confidence intervals always failed to cover the true RR of 1.0. However, once adjusted for the corresponding gestational age, the average RR was null (RR = 1.00) with 94% coverage. The average RR for the adjusted quintiles ranged from 0.97 to 1.03, with 94–95% coverage. The overall *z*-score model resulted in an average RR close to the expected null (0.97), but had a low coverage rate (84%), indicating that the observed bias is real (as opposed to random simulation variation), or the standard errors are being underestimated. The *z*-score model parameterised as quintiles resulted in RR ranging from 0.64 with quintile 3 to 0.90 with quintile 5 and coverage probabilities from 63% to 94%.

Table 3 describes the results by study site. The unadjusted estimated RR ranged from 0.78 at site 1 to 0.91 at site 6. After adjustment for gestational age the RR ranged from 0.97 to 1.00 with over 93% coverage at all sites. The *z*-score model results were highly variable depending on study site with estimates close to the true value of 1.0 at most sites, with confidence interval coverage ranging from 0.70 to 0.94, but only five of the 12 sites exceeding 90% coverage. In addition, risk estimates at some sites were noticeably biased, with the risk of neonatal mortality per *z*-score unit ranging from a 34% decreased risk (site 1, RR = 0.66, coverage = 70%) to a 43% increased risk (site 6, RR = 1.43, coverage = 80%). Further investigation suggested that this bias was likely due to a failure to remove the residual correlation between gestational weight gain and gestational age.

In addition, risk estimates from the *z*-score model were noticeably less precise than those from the model with total weight gain adjusted for gestational age. Standard error estimates of the regression coefficient under the *z*-score method were 2 to 5 times higher than those from the gestational age adjusted model. Thus, even when the *z*-score method produces unbiased estimates of the relative risk, the loss in precision can be detrimental, particularly when compared to the high levels of precision available via straightforward adjustment of gestational age.

Discussion

We have described the complexity gestational weight gain measures by demonstrating that total weight gain is a composite variable of cumulative weight gain. While this may seem intuitive, in practice it has implications

Table 1. Sample characteristics by study site (n = 65 643)

Characteristic	Site											
	1 (n = 2369) %	2 (n = 9719) %	3 (n = 2755) %	4 (n = 2849) %	5 (n = 24 129) %	6 (n = 2711) %	7 (n = 3486) %	8 (n = 1649) %	9 (n = 1667) %	10 (n = 2735) %	11 (n = 7434) %	12 (n = 4140) %
Age[a]	27.4 (6.6)	32.2 (5.7)	28.2 (6.2)	24.7 (5.8)	27.0 (5.1)	27.4 (5.6)	27.0 (6.8)	24.6 (6.2)	26.1 (6.4)	25.2 (6.1)	27.2 (6.7)	25.8 (6.3)
Race-ethnicity												
White	57.1	59.3	59.9	62.1	56.1	81.5	68.9	46.0	43.4	68.7	11.7	5.1
Black	15.2	8.0	11.2	19.4	29.4	0.5	4.2	35.2	38.3	24.6	47.1	24.9
Hispanic	17.7	19.2	15.7	9.8	10.8	10.0	7.4	2.5	13.3	2.2	31.6	58.0
Asian/ Pacific Islander	4.5	2.5	10.8	6.9	2.0	1.8	15.3	3.4	3.0	2.8	4.3	1.7
Other/ Unknown	5.6	11.1	2.4	1.8	1.7	6.2	4.3	12.9	2.1	1.8	5.2	10.3
Nulliparous	46.7	57.7	55.3	41.1	43.5	40.2	43.9	53.6	45.9	44.5	52.9	47.8
Married	66.8	43.0	86.4	64.3	34.7	84.8	84.2	42.4	29.4	47.2	28.2	45.0
Insurance												
Private	55.5	80.8	37.6	69.6	28.8	79.0	93.7	65.0	24.7	43.3	31.9	5.9
Public/Self pay	27.7	18.5	0.3	30.4	64.7	21.0	5.8	31.9	71.7	56.4	68.2	33.3
Other/Unknown	16.8	0.6	62.1	0.0	6.5	0.0	0.5	3.2	3.6	0.3	0.0	60.9
Smoked prior	5.8	10.7	1.7	11.0	20.0	3.7	2.4	0.2	19.0	19.3	12.1	4.2
Gestational weight gain												
Total, kg[a]	15.8 (5.7)	15.6 (5.4)	15.6 (5.7)	15.8 (6.5)	15.0 (5.3)	14.3 (5.0)	15.8 (6.0)	15.2 (6.2)	16.2 (6.3)	15.2 (6.5)	14.6 (6.1)	15.2 (6.3)
z-score[a,b]	−0.01 (1.00)	−0.11 (0.96)	−0.06 (1.07)	−0.13 (1.20)	−0.23 (1.02)	−0.32 (0.94)	−0.09 (1.16)	−0.22 (1.16)	−0.02 (1.16)	−0.22 (1.21)	−0.23 (1.15)	−0.18 (1.19)
Gestational age at delivery, wks[a]	38.3 (1.9)	39.0 (1.6)	38.5 (1.8)	38. 8 (2.0)	38.7 (1.6)	38.8 (1.5)	38.6 (2.1)	38.6 (2.0)	38.5 (2.2)	38.6 (2.2)	37.9 (2.5)	38.1 (2.5)

[a]Values represent mean (standard deviation).
[b]z-score estimated from published charts. Hutcheon *et al.*[3]

Table 2. Overall simulation results: total gestational weight gain and the risk for neonatal mortality estimated using a continuous measure and quintiles of total weight gain, unadjusted and adjusted for gestational age at delivery and using the weight-gain-for-gestational-age *z*-score

	GWG			GWG adjusted for GA			*z*-score		
Model	Mean RR[a]	Standard error	95% CI coverage	Mean RR[a]	Standard error	95% CI coverage	Mean RR[a]	Standard error	95% CI coverage
Linear[b]	0.87	0.02	0%	1.00	0.02	94%	0.97	0.08	84%
Quintiles[c]									
1	Reference			Reference			Reference		
2	0.30	0.26	0%	1.03	0.28	94%	0.78	0.26	84%
3	0.24	0.29	0%	1.02	0.30	95%	0.64	0.28	63%
4	0.21	0.32	0%	0.97	0.33	95%	0.76	0.26	83%
5	0.20	0.33	0%	1.01	0.34	95%	0.90	0.24	94%

CI, confidence interval; GA, gestational age at delivery; GWG, gestational weight gain; RR, mean relative risk across 5000 replicates.
[a]Expected RR based on simulations was 1.0. Simulations were repeated 5000 times.
[b]Linear estimates for GWG are per kg increase in weight gain, while *z*-score estimates are per *z*-score unit increase.
[c]Gestational weight gain quintile cut points (kg): 1, −9.1–10.9; 2, 10.9–13.6; 3, 13.6–15.9; 4, 15.9–19.5; 5, 19.5–49.0.

Table 3. Site specific simulation results: total gestational weight gain (GWG) and the risk for neonatal mortality estimated using a continuous measure of total weight gain, unadjusted and adjusted for gestational age at delivery and using the weight-gain-for-gestational-age *z*-score

		GWG			GWG adjusted for GA			*z*-score		
Site	*n*	Mean RR[a]	Standard error	95% CI coverage	Mean RR[a]	Standard error	95% CI coverage	Mean RR[a]	Standard error	95% CI coverage
1	2369	0.78	0.27	47%	0.98	0.18	95%	0.66	0.59	70%
2	9719	0.89	0.08	44%	1.00	0.06	96%	1.02	0.31	92%
3	2755	0.87	0.12	55%	0.98	0.22	94%	0.93	0.50	73%
4	2849	0.87	0.09	47%	0.99	0.11	95%	0.87	0.32	84%
5	24 129	0.88	0.05	12%	1.00	0.04	94%	0.96	0.13	86%
6	2711	0.91	0.17	89%	0.98	0.18	95%	1.43	0.72	80%
7	3486	0.87	0.07	35%	0.99	0.07	94%	1.03	0.23	93%
8	1649	0.89	0.11	75%	0.97	0.16	95%	1.03	0.44	92%
9	1667	0.90	0.08	70%	0.98	0.15	94%	1.05	0.42	90%
10	2735	0.90	0.09	61%	0.99	0.08	93%	1.01	0.39	87%
11	7434	0.88	0.04	8%	1.00	0.04	94%	1.10	0.18	88%
12	4140	0.88	0.05	16%	1.00	0.05	95%	1.01	0.21	94%

CI, confidence interval; GA, gestational age at delivery; GWG, gestational weight gain; RR, mean relative risk across 5000 replicates.
[a]Estimates for GWG are per kg increase in weight gain, while *z*-score estimates are per *z*-score unit increase. Expected RR based on simulations is 1.0. Simulations were repeated 5000 times.

for how we implement this variable in epidemiological models. From the DAG presented, it is clear that total gestational weight gain is a time-varying exposure, although it is often treated as a fixed variable. We show that although there is feedback between gestational age and weight gain across time, when limited to studies of total gestational weight gain and assessing the association with neonatal mortality, adjusting for gestational age blocks the open backdoor path between these prior longitudinal relationships. This finding is supported by our simulated model of no true association between weight gain and neonatal mortality, where we demonstrated that efficient unbiased estimates can be achieved by adjusting for gestational age.

Adjusting for gestational age at delivery in studies of total gestational weight gain has previously been of debate.[5–7] Using simulated data, we found that when

adjusted for gestational age, total weight gain can be used to obtain unbiased estimates of the true association with neonatal mortality assuming no unmeasured confounding. The average rate of weight gain, calculated as total weight gain divided by the length of gestation, has been previously used to partially account for the time at risk; however, this approach also produces biased results.[2] The rate of weight gain method partially corrects for gestational length, but ratio measures require strong assumptions and therefore regression-based approaches are preferred.[13] Using this straightforward adjustment-based approach, we were able to break the link between weight gain and gestational age. This adjustment directly permits investigators to test various functional forms of this variable in relation to the outcome. Another advantage is that the model permits flexibility in identifying the most appropriate relationship between potential confounders, such as race or socioeconomic status, with the exposure and outcome.

Recently, a weight gain for gestational age *z*-score was proposed as a method to address the bias in gestational weight gain analyses due to the inherent correlation with gestational age.[3] While this method likely performs well when correctly specified, such that the *z*-scores adequately represent the underlying population, our findings suggest that violations in these assumptions, such as applying *z*-score derived in one population to another, can result in *z*-score estimates that are no longer independent of gestational age, retaining residual confounding of gestational age, and subsequently resulting in biased estimates. In additional simulations (not shown here), we verified that when assumptions are appropriately met, both the *z*-score model and the adjusted model produce unbiased estimates, but the *z*-score model is subject to considerably reduced precision. We leveraged existing data from a multi-site cohort to evaluate how changing the underlying demographics of the sample impacted the ability of the *z*-score to fully remove confounding due to gestational age. Overall the *z*-score method produced biased results of the estimated RR and its standard error. Although the bias of the RR estimate at some of the sites appeared minimal, the low CI coverage rates suggest that either the bias is non-negligible, or the *z*-score model is underestimating the standard error, or both. Theoretically, if the study sample for a given site were similar to the population from which the *z*-scores were created, estimates from the *z*-score model would be unbiased; however, this property is difficult to ascertain given only the observed study sample. We hypothesise that the observed bias was most likely due to misspecification of the gestational weight gain and gestational age relationship, which led to incompatibility of the *z*-score when applying it to a sample outside of the population used for the reference charts. It is apparent that the weight gain distribution in our cohort different from the reference population as the average *z*-score varied from -0.01 to -0.32 depending on study site, indicating that there is large variation in women's weight gain distribution across the United States. Notably, conformance to a mean of 0 and variance of 1 does not necessarily guarantee against misspecification. For example, at site 1 the average *z*-score was -0.01 with a standard deviation of 1.00, but the resulting relative risk estimate was biased by 34%. Although we did not evaluate the site-specific results by quintiles, due to the small number of cases by site, the overall quintile simulation results showed even greater bias than the linear models. Therefore, researchers should be cautious when applying the published *z*-score charts to their own data as it may not be properly specified to adequately remove confounding due to gestational age. As such, it is important to recognise that the bias estimates provided in this analysis are specific to the study population at hand and should not be applied as a means to correct bias in other studies.

It has been suggested that nationally representative charts for gestational weight gain are needed.[3,4] Similar to birthweight *z*-scores,[14] if developed, nationally representative charts would be useful for comparing weight gain trajectories across populations or trends in total weight gain[15] and would have the advantage of standardising for changes in gestational age between populations or in time. As demonstrated in our analyses, however, the *z*-score may not adequately remove the bias due to confounding from gestational age in epidemiological studies, when utilised to relate total weight gain and pregnancy outcomes.

When respective assumptions are correctly specified, the regression-based approach and the *z*-score both yield unbiased estimates. The *z*-score method, however, implicitly relies on correct specification of the relationship between gestational age and weight gain used to calculate the values in the *z*-score chart. These assumptions cannot be validated using only the data on hand, and misspecification of the *z*-score model can result in *z*-score values that are not

independent of gestational age. The regression-based approach, however, has a more flexible framework to test the assumed relationship between gestational age and weight gain, and does not rely on the correct estimation of additional parameters as well as the commutability of an external data set (via the *z*-score chart), to the study population. We have provided the algebraic proofs for these concepts in Appendix A.

The interpretation and use of anthropometric *z*-scores has been an important area of discussion for many years.[14,16] Z-score reference charts maybe useful for comparing maternal weight gain across populations, just as is done with child weight-for-age *z*-scores; however, extensive research would be required to determine optimal *z*-score cut points associated with healthy outcomes before being translated into practice and making clinical decisions. Furthermore, using the *z*-score charts in clinical practice would require that women use *z*-score charts to monitor their weight gain, as opposed to tracking their weight in pounds, which may add an additional layer of complexity to helping women achieve a healthy weight during pregnancy. As we have shown that for epidemiological analyses the *z*-score must be specific to the source population, *z*-scores may be incomparable across studies, further complicating the public health interpretation.

Some limitations of this research warrant discussion. Our simulations only covered the scenario of a null association between weight gain and mortality, and were not intended to provide a comprehensive study of all possible scenarios. Theoretically, so long as the DAG holds, the models should demonstrate similar results under other scenarios, specifically, the *z*-score model will be biased when applied to a study sample that differs fundamentally from the population used to generate the *z*-scores. The magnitude of the bias and confidence interval coverage rates, however, will vary depending on the true association between weight gain and mortality. Also, in Figure 1, we made the assumption of no measured or unmeasured confounding variables. If unmeasured confounding of the gestational age-mortality relation is present, neither the proposed regression-based method nor the *z*-score can avoid collider bias,[10] and more advanced methods such as marginal structural models or g-estimation may be necessary.[17] Therefore, all relevant measured confounding variables should be considered in the model with the goal of reducing unmeasured confounding. Another important outcome clearly related to gestational age at delivery is preterm delivery. Preterm delivery (gestational age <37 weeks) as defined today, is a deterministic function of gestational age at delivery. Thus, when investigating preterm delivery as an outcome, adjusting for gestational age as we did in our models of neonatal mortality can result in non-convergence due to overfitting. We have shown previously that models designed for time to event variables can successfully overcome these issues.[18] Many outcomes (e.g. cognitive development) related to gestational weight gain can benefit from the proposed adjustment method, but the specific DAG should be considered for each question and no single approach fits all situations.

We have shown that total gestational weight gain can be used to achieve unbiased results when the outcome is associated with gestational age by simply adjusting for gestational age at delivery. Furthermore, applying *z*-score charts to external studies outside of the underlying population can lead to misspecification, and even when assumptions are appropriately met, the *z*-score model is subject to reduced precision. Thus, because of the added complexity of calculation and interpretation, as well as the substantial efficiency loss, the utility of the *z*-score is unclear and caution is warranted before they are implemented in epidemiological studies. It is important to note that neither total weight gain nor the *z*-score allow for the examination of critical windows in gestation, when the level of weight gain may be more or less important for the respective outcome. Similarly, both methods assume that rank preservation and women continue on the same trajectory of weight gain across gestation relative to their peers. Therefore, serial measures of weight gain remain optimal. Emerging research suggests that gestational weight gain may be modifiable and therefore a potential target for reducing risk of adverse outcomes.[19] Thus, achieving unbiased estimates of the relation between weight gain and pregnancy outcomes is important for improving maternal and child health.

Acknowledgements

This research was supported by the Intramural Research Program of the National Institutes of Health, *Eunice Kennedy Shriver* National Institute of Child Health and Human Development (contract number

HHSN267200603425C). The authors would like to thank Drs. Nansi S. Boghossian, Jennifer A. Hutcheon, Robert W. Platt and Lindsey A. Sjaarda for their helpful suggestions to earlier versions of the manuscript. The authors have no conflicts of interest.

References

1 Institute of Medicine and National Research Council. *Weight Gain During Pregnancy: Reexamining the Guidelines*. Washington, DC: National Academies Press, 2009.
2 Hutcheon JA, Bodnar LM, Joseph KS, Abrams B, Simhan HN, Platt RW. The bias in current measures of gestational weight gain. *Paediatric and Perinatal Epidemiology* 2012; 26:109–16.
3 Hutcheon JA, Platt RW, Abrams B, Himes KP, Simhan HN, Bodnar LM. A weight-gain-for-gestational-age z score chart for the assessment of maternal weight gain in pregnancy. *The American Journal of Clinical Nutrition* 2013; 97:1062–7.
4 Hutcheon JA, Platt RW, Abrams B, Himes KP, Simhan HN, Bodnar LM. Pregnancy weight gain charts for obese and overweight women. *Obesity (Silver Spring)* 2015; 23:532–5.
5 Davis RR, Hofferth SL, Shenassa ED. Gestational weight gain and risk of infant death in the United States. *American Journal of Public Health* 2014; 104:S90–5.
6 Davis RR, Hofferth SL, Shenassa ED. Davis et al. respond. *American Journal of Public Health* 2014;104:e2.
7 Hutcheon JA, Bodnar LM, Abrams B. Untangling gestational weight gain from gestational age in infant mortality studies. *American Journal of Public Health* 2014; 104:e1–2.
8 VanderWeele TJ, Mumford SL, Schisterman EF. Conditioning on intermediates in perinatal epidemiology. *Epidemiology* 2012; 23:1–9.
9 Greenland S. Quantifying biases in causal models: classical confounding vs collider-stratification bias. *Epidemiology* 2003; 14:300–6.
10 Whitcomb BW, Schisterman EF, Perkins NJ, Platt RW. Quantification of collider-stratification bias and the birthweight paradox. *Paediatric and Perinatal Epidemiology* 2009; 23:394–402.
11 Zhang J, Troendle J, Reddy UM, Laughon SK, Branch DW, Burkman R, *et al*. Contemporary cesarean delivery practice in the United States. *American Journal of Obstetrics and Gynecology* 2010;203:326.e1-.e10.
12 Rothman KJ, Greenland S, Lash TL. *Modern Epidemiology*. Philadelphia, PA: Wolters Kluwer Health/Lippincott Williams & Wilkins, 2008.
13 Allison DB, Paultre F, Goran MI, Poehlman ET, Heymsfield SB. Statistical considerations regarding the use of ratios to adjust data. *International Journal of Obesity and Related Metabolic Disorders* 1995; 19:644–52.
14 Delbaere I, Vansteelandt S, De Bacquer D, Verstraelen H, Gerris J, De Sutter P, *et al*. Should we adjust for gestational age when analysing birth weights? The use of z-scores revisited. *Human Reproduction* 2007; 22:2080–3.
15 Johnson JL, Farr SL, Dietz PM, Sharma AJ, Barfield WD, Robbins CL. Trends in gestational weight gain: the Pregnancy Risk Assessment Monitoring System, 2000-2009. *American Journal of Obstetrics and Gynecology* 2015; 28:00086–1.
16 Dibley MJ, Staehling N, Nieburg P, Trowbridge FL. Interpretation of Z-score anthropometric indicators derived from the international growth reference. *American Journal of Clinical Nutrition* 1987; 46:749–62.
17 Robins JM, Hernan MA, Brumback B. Marginal structural models and causal inference in epidemiology. *Epidemiology* 2000; 11:550–60.
18 Mitchell EM, Hinkle SN, Schisterm EF. It's About Time: a survival approach to gestational weight gain and preterm delivery. *Epidemiology* 2016; 27:182–7.
19 Hill B, Skouteris H, Fuller-Tyszkiewicz M. Interventions designed to limit gestational weight gain: a systematic review of theory and meta-analysis of intervention components. *Obesity Reviews* 2013; 14:435–50.

Appendix A

Consider a z-score chart where z-scores are calculated by relating gestational weight gain (GWG) during pregnancy to gestational age (t) through a linear model such that:

$$\mathrm{GWG}(t) = \psi_0 + \psi_1 t + \psi_2 t^2 + \varepsilon_1, \qquad 1$$

where the ε_1's are identically distributed with mean 0 and variance σ^2. The variable t is chosen to represent gestational age to emphasise the equivalence of gestational age with time, and GWG(t) denotes that GWG is a function of time. Then the z-score calculated from this model is equivalent to the standardised residuals:

$$Z = \frac{\mathrm{GWG}(t) - \hat{E}(\mathrm{GWG}|t)}{\sqrt{\widehat{\mathrm{Var}}(\mathrm{GWG}|t)}} = \frac{\varepsilon}{\sigma}.$$

The goal of using a z-score is to provide a marker of relative gestational weight gain that is independent of gestational age. This method will be successful so long as the assumptions in (1) are met, namely, that ε_1 is independent of gestational age with mean 0.

Suppose, however, that model (1) is misspecified, and in fact, the true model for GWG includes an interaction between gestational age (t) and pre-pregnancy body mass index (BMI), such that:

$$\mathrm{GWG}(t) = \psi_0 + \psi_1 t + \psi_2 t^2 + \psi_3(\mathrm{BMI} * t) + \varepsilon_2.$$

Then fitting model (1) will violate the assumption of independence between the errors and t, since the error terms are now inclusive of the interaction term:

$$\varepsilon_1 = \psi_3(\mathrm{BMI} * t) + \varepsilon_2.$$

which is clearly correlated with t. Thus, z-scores calculated based on fitting model (1) will not succeed in removing the association between the z-score and gestational age. Subsequently, z-scores based on total gestational weight gain and gestational age at delivery will fail to adequately remove these correlations. This performance is not only hindered by the omission of interaction terms of gestational age, but could also be enforced by failing to include a variable that acts as a confounder between gestational age and GWG, or between gestational age and the outcomes of interests (e.g. neonatal mortality). For instance, if gestational age is caused by some unmeasured variable C, failure to incorporate C into Model (1) will result in a z-score that remains correlated with C. Subsequently, if C is a confounder of the relationship between gestational age and neonatal mortality, C will remain a confounder of the z-score and neonatal mortality. Essentially, the model used to calculate the z-score chart must correctly specify the relationship between gestational age and GWG in order for the z-scores to perform as expected.

Now consider the logistic regression model of interest relating the z-score to the binary outcome of interest (Y), where total GWG ($\mathrm{GWG_{tot}}$) is measured at delivery ($t = \mathrm{GA_{del}}$), and:

$$\begin{aligned}&\mathrm{logit}[\Pr(Y = 1|Z)]\\ &\quad = \alpha + Z\beta = \alpha + \left(\frac{\mathrm{GWG}_{tot} - \hat{E}(\mathrm{GWG}_{tot}|\mathrm{GA}_{del})}{\sqrt{\widehat{\mathrm{Var}}(\mathrm{GWG}_{tot}|\mathrm{GA}_{del})}}\right)\beta.\end{aligned}$$

For simplicity, let us suppose that $\mathrm{Var}(\mathrm{GWG_{tot}}|\mathrm{GA_{del}}) = 1$ at all gestational ages to retain the original scale for the log-odds (β). Then

$$\begin{aligned}\mathrm{logit}[\Pr(Y = 1|Z)] =& \alpha + Z\beta = \alpha + (\mathrm{GWG}_{tot})\beta - (\hat{\psi}_0\beta)\\ &- (\hat{\psi}_1\beta)\mathrm{GA}_{del} - (\hat{\psi}_2\beta)\mathrm{GA}^2_{del}\\ &= \alpha^* + (\mathrm{GWG}_{tot})\beta + \beta_1^*\mathrm{GA}_{del}\\ &+ \beta_2^*\mathrm{GA}^2_{del}\end{aligned}.$$

Note that this model can be accomplished by fitting a logistic regression with total GWG as the exposure of interest, with gestational age at delivery ($\mathrm{GA_{del}}$) and $\mathrm{GA^2_{del}}$ as covariates. In addition, a direct model permits more flexibility in choosing the functional form of the covariates (e.g. $\mathrm{GA_{del}}$ and $\mathrm{GA^2_{del}}$), or additional terms (e.g. BMI * $\mathrm{GA_{del}}$). Furthermore, it does not restrict estimation of β by pre-specifying the relationship between gestational weight gain and gestational age, as defined by Model (1) and its estimated parameters (ψ). Application of z-scores generated under a misspecified model can result in biased inference that is untestable using only the study sample. In conclusion, while the z-scores will theoretically perform well under correct specification, it may be preferable to limit the required assumptions to those that can be tested via regression models on the study sample.