

A systematic approach for establishing the range of recommended weight gain in pregnancy^{1–3}

Jennifer A Hutcheon and Lisa M Bodnar

ABSTRACT

Background: Current approaches for establishing public health guidelines on the recommended range of weight gain in pregnancy are subjective and nonsystematic.

Objective: In this article, we outline how decision-making on gestational weight-gain guidelines could be aided by quantitative approaches used in noninferiority trials.

Design: We reviewed the theoretical application of noninferiority margins to pregnancy weight-gain guidelines. A worked example illustrated the selection of the recommended range of pregnancy weight gain in women who delivered at the Magee-Womens Hospital, Pittsburgh, PA, in 2003–2010 by identifying weight-gain z scores in which risk of unplanned cesarean delivery, preterm birth, small-for-gestational-age infant, and large-for-gestational-age infant were not meaningfully increased (based on noninferiority margins of 10% and 20%).

Results: In normal-weight women, lowest risk of adverse perinatal outcome was observed at a weight-gain z score of -0.2 SDs. With a noninferiority margin of 20%, risks of adverse outcome were not meaningfully increased from the -0.2 -SD reference value between z scores of -0.97 and $+0.33$ SDs (which corresponded to 11.3–18.4 kg). In overweight women, the recommended range was much broader: -2.11 to $+0.29$ SDs (4.4–18.1 kg).

Conclusion: The new approach illustrated in this article has a number of advantages over current methods for establishing pregnancy weight-gain guidelines because it is systematic, it is reproducible, and it provides a tool for policy makers to derive guidelines that explicitly reflect values at which risk of adverse outcome becomes meaningfully increased. *Am J Clin Nutr* 2014;100:701–7.

INTRODUCTION

In 2009, the Institute of Medicine (IOM) published revised guidelines on how much weight women should gain during pregnancy (1). The recommendations attempt to optimize a variety of short- and long-term maternal and child health outcomes including fetal growth, preterm birth, and maternal and offspring obesity (1). For women of normal prepregnancy BMI (in kg/m^2 ; 18.5–24.9), the recommended range of total weight gain is 11.5–16 kg, whereas for underweight (<18.5), overweight (25–29.9), and obese (≥ 30) women, ranges are 12.5–18, 7–11.5, and 5–9 kg, respectively. In selecting these ranges, the IOM Committee relied heavily on observational research from several populations that presented plots of risks of adverse outcomes across categories of gestational weight gain. The IOM Committee did not have a systematic, reproducible process through which to

establish weight-gain thresholds at which risks become unacceptably increased. Instead, they visually inspected curves to estimate the range of weight gain within which risks remained low (ie, “eyeballing” optimal cutoffs) (1). As a result, the recommended ranges are subjective and may be either excessively narrow (leading to unnecessary clinical intervention to correct the course of weight gain) or excessively wide (failing to identify women who would benefit from clinical intervention to optimize weight gain).

We propose that decision-making on gestational weight-gain guidelines could be aided by quantitative approaches used in noninferiority trials. Noninferiority trials are a type of randomized clinical trial in which the goal is to determine whether a new intervention is at least as good as (not meaningfully worse than) a standard intervention (2). These trials are often performed to show that a new intervention with more desirable qualities (eg, lower costs, fewer side-effects, or less frequent dosages) is at least as effective as a current standard. For example, the WHO conducted a noninferiority trial in developing countries to test the hypothesis that a new model of antenatal care with fewer antenatal visits led to maternal and perinatal outcomes that were at least as good as the traditional care model (3). To establish noninferiority, the WHO team specified that risk of adverse maternal and perinatal outcomes could not be $>20\%$ higher with the new model. This difference in risk, which is known as the noninferiority margin, reflects the prespecified point at which patients and caregivers believe the 2 treatments can no longer be considered comparable. In the current study, our objective was to illustrate how the use of noninferiority margins can provide

¹ From the Department of Obstetrics and Gynaecology, University of British Columbia, Vancouver, Canada (JAH), and the Departments of Epidemiology and Obstetrics, Gynecology, and Reproductive Sciences, Graduate School of Public Health and School of Medicine, University of Pittsburgh, Pittsburgh, PA (LMB).

² Supported by the Eunice Kennedy Shriver National Institute of Child Health and Human Development (R01 HD072008; to LMB and JAH). JAH holds a Canadian Institutes of Health Research New Investigator Award and is a Career Scholar of the Michael Smith Foundation for Health Research.

³ Address reprint requests and correspondence to J Hutcheon, Department of Obstetrics and Gynaecology, University of British Columbia, Shaughnessy Room C408A, BC Women’s Hospital & Health Centre, 4500 Oak Street, Vancouver, British Columbia V6N 1H3, Canada. E-mail: jhutcheon@cfri.ca.

Received February 4, 2014. Accepted for publication May 21, 2014.

First published online June 25, 2014; doi: 10.3945/ajcn.114.085258.

a quantitative approach for establishing the healthiest range of recommended weight gain in pregnancy.

SUBJECTS AND METHODS

Review of noninferiority margins

In a noninferiority trial, the noninferiority margin plays a central role in the decision of whether a new intervention can be concluded to be “not meaningfully worse” than the standard treatment (2, 4, 5). The noninferiority margin is typically selected by using expert clinical opinion and/or statistical limits. Investigators may conduct a survey of health care providers for their views on the minimum difference they consider to be clinically important in risk between groups (6). Some regulatory agencies have published recommended values for noninferiority margins [eg, a margin of 10% is recommended by the Committee for Health and Medicinal Products, whereas the Federal Drug Administration has previously recommended margins of 10%, 15%, and 20% depending on the prevalence of the adverse outcome in the control group (4)]. The margin can be expressed in either absolute terms (eg, a risk difference of no more than 5 per 100) or relative terms (eg, an increase in risk of no more than 15%).

After completion of the trial, risk of the primary outcome is compared between the 2 treatment groups, either by calculating a risk difference or an RR with its 95% CI. The range of the 95% CI is then compared with the noninferiority margin. To conclude that a new intervention is not meaningfully worse than a standard intervention (ie, noninferior), the 95% CI must be entirely below the noninferiority margin. Conclusions on noninferiority compared with inferiority reflect the clinical relevance of any observed differences between groups.

Three possible results from a hypothetical noninferiority trial with a prespecified noninferiority margin of 15% (RR: 1.15) are shown in **Figure 1**. In scenario A, the RR for the comparison of the new intervention with the standard intervention was 1.04 (95% CI: 0.99, 1.09). Because the upper limit of the 95% CI (1.09) was below the margin of 1.15, it could be concluded that the new intervention was noninferior to the standard intervention. In other words, the new intervention was at least as effective as the standard. In scenario B, the RR was 1.22 (95% CI: 1.17, 1.27). Because the 95% CI was entirely above the noninferiority margin of 1.15, it could be concluded that the new intervention led to worse outcomes than the standard in-

tervention. In scenario C, the RR of 1.14 was below the noninferiority margin, but its CI reached an upper limit of 1.19, thereby crossing the noninferiority margin. Therefore, the study was inconclusive because the range of the 95% CI was compatible with the conclusion that the new intervention was worse than the standard intervention (eg, the 95% CI was compatible with a true RR of 1.17, which would be considered to be worse than the standard intervention) but was also compatible with the conclusion that the new intervention was no worse than the standard intervention (eg, the 95% CI was compatible with a true RR of 1.09, which would be considered to be clinically equivalent to the standard intervention). Neither possibility can be ruled out, and thus, conclusions on noninferiority or inferiority cannot be made.

Application of noninferiority margins to selection of gestational weight-gain ranges

Noninferiority margins are traditionally used in clinical trials to determine whether the risk associated with a new treatment is not meaningfully higher than the risk associated with a standard treatment. However, we propose that noninferiority margins can also be used with observational data to identify points on the gestational weight-gain continuum where risks become meaningfully increased (ie, noninferior) compared with the nadir of risk (ie, the gestational weight gain at which risks of adverse outcomes are lowest). A prespecified, reproducible decision-making strategy such as this would ensure that the cutoffs for recommended ranges reflect weight gains above or below which risks are deemed to be unacceptably increased.

This approach can be implemented through the following steps:

1) Establish the noninferiority margin on the basis of input from health policy makers, health care providers, pregnant women, and other key stakeholders in creating pregnancy weight-gain guidelines (6). For example, stakeholders may agree that an increase in childhood obesity risk of 9% or less is not clinically meaningful but that increased risks of 10% or greater can no longer be considered equivalent. This method would result in a 10% noninferiority margin. The process of establishing noninferiority margins should take into account the different baseline risks between groups (eg, obese compared with normal-weight women), and different margins of noninferiority could potentially be chosen for different groups.

2) Create a multivariable regression model that describes risks of adverse pregnancy outcomes as a function of (continuous) gestational weight gain. Because the relation between gestational weight gain and adverse pregnancy outcomes has been shown to follow a nonlinear shape in risk (1), a flexible, nonlinear approach should be used to model gestational weight gain, such as a restricted cubic spline (7). The model should also be adjusted for relevant confounders. Care should be taken to ensure that the relation between weight gain and the adverse outcome is modeled accurately. Different knot positions and numbers of knots should be assessed to identify the best-fitting model by using criteria such as the Akaike information criterion or the Bayesian information criterion.

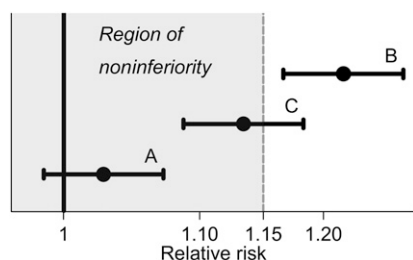


FIGURE 1. Hypothetical results (RRs with 95% CIs) from a noninferiority trial. Conclusions on noninferiority are based on the range of the 95% CI (horizontal lines with caps) in relation to a prespecified margin of noninferiority indicated by the vertical dashed line (in this example, a 15% increase or an RR of 1.15). Noninferiority is established when the 95% CI is entirely below the noninferiority margin (line A). Results are inconclusive if the CI crosses the margin of noninferiority (line C), whereas the new treatment is inferior if the CI is entirely above the margin of noninferiority (line B).

3) Use the regression equation estimated by the model in step 2 to obtain the adjusted predicted probabilities of an adverse pregnancy outcome across the gestational weight-gain continuum. Plot the predicted probabilities and identify the gestational weight gain value at which the predicted probability of adverse outcome is lowest (the nadir of risk); this weight gain value will serve as the referent.

4) Calculate RRs (or risk differences) and 95% CIs for comparison of risk of weight gains above and below the referent weight gain with risk at the referent. Users can readily calculate these measures of effect by using post-estimation commands such as `.margins` or `.xbli` in the Stata program (StataCorp) or with the effects package in R software (Project CRAN).

5) Identify weight-gain values at which the upper limit of the 95% CI exceeds the prespecified margin of noninferiority. If there is a U shape in risk (increased risks associated with both high and low weight gains), there will be a point above and a point below the nadir where risks exceed the noninferiority margin. At these values, risks of adverse outcome can no longer be considered to be the same as risks at the nadir and, therefore, represent appropriate cutoffs for gestational weight-gain range recommendations.

Worked example

We illustrate the application of this approach in a cohort of women delivering at the Magee-Womens Hospital in Pittsburgh, PA, from 2003 to 2010. Data were obtained from a comprehensive perinatal database on all deliveries at the institution. Data came from admitting services, medical record coding (procedure and diagnosis codes), medical record abstraction, the electronic birth record, and other ancillary systems. Fields with missing data were completed by using information from the birth certificate when possible. Personal identifying information in the database was eliminated to ensure confidentiality, and the University of Pittsburgh Institutional Review Board approved the study.

We included all singleton, live-born infants from 20 to 42 wk of gestation with complete data on prepregnancy weight, height, maternal weight at delivery, and key covariates. Prepregnancy weight and height were ascertained based on self-report at the first prenatal visit. Maternal weight at delivery was gathered from the last measured prenatal weight. Prepregnancy BMI was calculated as prepregnancy weight divided by height squared. For simplicity, we restricted our analyses to 2 prepregnancy BMI groups of normal weight and overweight women. Gestational weight gain was calculated as the difference between maternal weight at delivery and prepregnancy weight. Gestational age-standardized maternal weight-gain *z* scores were assigned on the basis of BMI-specific percentile charts developed in this population (8).

Gestational weight gain is associated with increased risks of a number of different health outcomes, and thus, we created a composite of adverse maternal and neonatal health outcomes. Our composite outcome included health outcomes identified by the IOM committee as having a potentially important link with gestational weight gain and for which we had available data. Our composite included any of the following: small-for-gestational-

age infant, large-for-gestational-age infant, unplanned cesarean delivery, spontaneous preterm birth, and indicated preterm birth. We defined spontaneous preterm birth as a delivery of a live infant at 20 to <37 completed weeks of gestation after preterm labor with intact membranes or a preterm prelabor rupture of fetal membranes. All remaining preterm births were classified as indicated. Small- and large-for-gestational-age births were defined as live-born infants who were less than the 10th percentile or greater than the 90th percentile, respectively, of ultrasound-based intrauterine fetal weight standards (9). An unplanned cesarean delivery was defined as a cesarean delivery after labor. In the absence of data or agreement on how to weight outcomes relative to one another according to their health impacts (1), components of composite outcomes were weighted equally.

Because, to our knowledge, there are no published studies that have elicited expert opinion on noninferiority margins for gestational weight gain, we considered 2 different statistical margins of noninferiority at 10% and 20%. As detailed in steps 2–4, we estimated the relation between gestational weight gain [modeled as a restricted cubic spline with 5 knots in the default positions (7)] and the composite adverse outcome by using log-binomial regression. We built separate models for normal-weight and overweight women. Models were adjusted for the year of birth, race-ethnicity, parity, marital status, education, insurance status at delivery, smoking, maternal age, and height. We established the weight-gain value at which risk of an adverse outcome was lowest and calculated the RR of adverse outcome at all other gestational weight-gain values (in increments of 0.1 *z* scores). Finally, we identified weight-gain *z* scores at which the upper limit of the 95% CI for the RR (for comparison of risk of adverse outcome at that weight gain with risk at the nadir) was just below 1.10 (for a 10% noninferiority margin) or 1.20 (for a 20% noninferiority margin). Analyses were conducted with Stata version 12 software.

RESULTS

There were 68,532 deliveries of singleton live-born infants from 20 to 42 wk of gestation at Magee-Womens Hospital from 2003 to 2010. With the exclusion of 19,164 (28%) deliveries that lacked information on prepregnancy weight, height, or maternal weight at delivery or covariates in the final model, 49,368 deliveries were left in our study population. Our analyses were based on the 27,585 normal-weight (BMI: from 18.5 to <25) and 10,989 overweight (BMI: from 25 to <29) (10) women in the cohort. These women were predominantly non-Hispanic white (78%), college educated (51%), and married (64%). The mean (\pm SD) maternal age was 29.1 ± 6.1 y. Approximately one-half (47%) of the women were nulliparous, and 15% of the women smoked.

The pregnancy weight gain and risk of each adverse outcome in normal-weight and overweight women are shown in **Table 1**. Normal-weight women had 32.2% risk of having a delivery with at least one of the components of the composite adverse outcome, whereas in overweight women, this risk was 37.2%. The most-commonly occurring components of the composite outcome were an unplanned cesarean delivery and small for gestational age in normal-weight women and unplanned cesarean delivery and large for gestational age in overweight women.

TABLE 1
Risks of adverse pregnancy outcome by gestational weight-gain *z* score in 38,574 singleton births to normal weight or overweight women at the Magee-Womens Hospital in Pittsburgh, PA, 2003–2010¹

	All women	Gestational weight-gain <i>z</i> score		
		Less than −1 SD	−1 to +1 SD	Greater than +1 SD
Normal weight (<i>n</i>)	27,585	3649	21,328	2608
Weight gain (kg)	15.9 ± 6.0 ²	7.2 ± 3.1	15.9 ± 3.5	27.6 ± 5.8
SGA [<i>n</i> (%)]	2557 (9.3)	656 (18.0)	1756 (8.2)	145 (5.6)
LGA [<i>n</i> (%)]	2009 (7.3)	93 (2.6)	1514 (7.1)	402 (15.4)
Unplanned cesarean [<i>n</i> (%)]	3445 (13.8)	369 (11.1)	2620 (13.6)	455 (19.6)
Spontaneous preterm birth [<i>n</i> (%)]	1407 (5.3)	275 (8.0)	985 (4.8)	147 (5.9)
Iatrogenic preterm birth [<i>n</i> (%)]	983 (3.8)	198 (5.9)	657 (3.2)	128 (5.2)
Any adverse outcome [<i>n</i> (%)]	8874 (32.2)	1347 (36.9)	6491 (30.4)	1035 (39.7)
Overweight (<i>n</i>)	10,989	1132	9347	510
Weight gain (kg)	15.2 ± 7.4	3.0 ± 4.1	15.7 ± 5.0	32.4 ± 6.1
SGA [<i>n</i> (%)]	967 (8.8)	184 (16.3)	745 (8.0)	38 (7.5)
LGA [<i>n</i> (%)]	1180 (10.7)	56 (5.0)	1020 (10.9)	104 (20.4)
Unplanned cesarean [<i>n</i> (%)]	1713 (17.9)	108 (10.7)	1495 (18.4)	110 (25.9)
Spontaneous preterm birth [<i>n</i> (%)]	557 (5.3)	77 (7.2)	458 (5.1)	22 (4.9)
Iatrogenic preterm birth [<i>n</i> (%)]	506 (4.9)	67 (6.4)	380 (4.3)	59 (12.1)
Any adverse outcome [<i>n</i> (%)]	4092 (37.2)	415 (36.7)	3419 (36.6)	258 (50.6)

¹ LGA, large-for-gestational-age infant (>90th percentile); SGA, small-for-gestational-age infant (<10th percentile).
² Mean ± SD (all such values).

For normal-weight women, the relation between gestational weight-gain *z* scores and individual adverse outcomes that made up the composite is shown in **Figure 2A**. The graph highlights that higher weight-gain *z* scores were linked with increased risks of large for gestational age and unplanned cesarean, whereas lower weight-gain *z* scores increased risk of small for gestational age. Risks of spontaneous and indicated preterm birth followed a slight U shape across gestational weight gain.

The association between gestational weight gain and risk of the composite adverse outcome in normal-weight women is shown in **Figure 3A**. As expected, a U shape in risk was observed, with increased risks at both lower and higher weight gains. Lowest risk of an adverse outcome was 28.3/100 and occurred at

a weight-gain *z* score of −0.2 (the equivalent of 15.3 kg at 40 wk of gestation). Risk of adverse outcome changed smoothly across the weight-gain continuum, and thus, there were no obvious cutoffs at which to establish upper and lower limits of recommended weight-gain ranges.

We calculated RRs of adverse outcome at each weight-gain *z* score by using risk at the nadir (28.3/100) as the reference value. See Supplemental Table 1 under “Supplemental data” in the online issue for these RRs with 95% CIs. At the lower end of the U shape, the 95% CI first included the 10% noninferiority margin at a *z* score of −0.59 (adjusted RR: 1.05; 95% CI: 1.01, 1.09) (**Table 2**). At the upper end of the U shape, the 95% CI remained within the 10% margin until a *z* score of +0.15

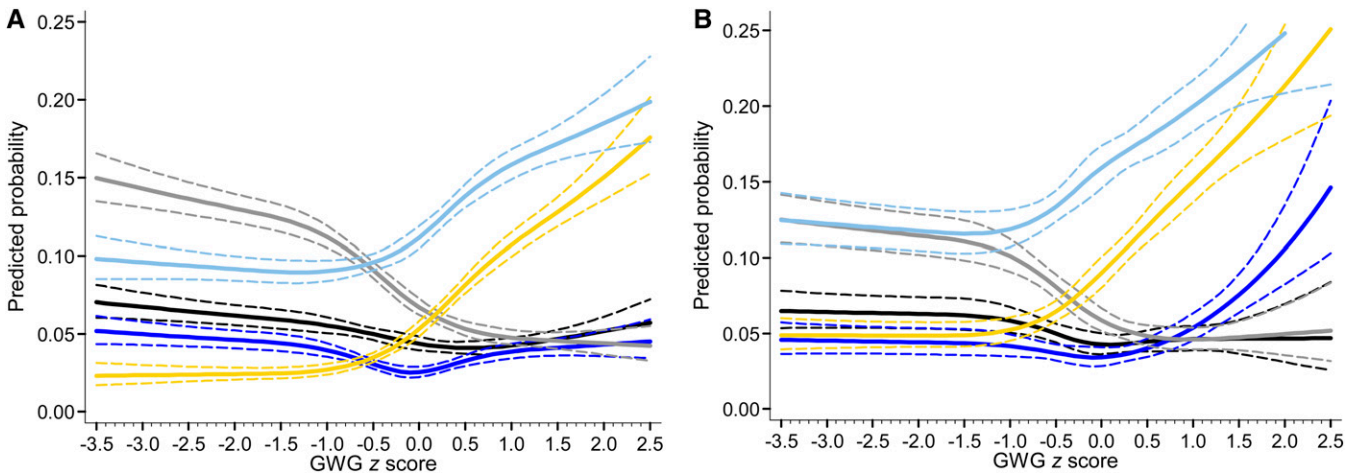


FIGURE 2. Adjusted predicted probability of 5 adverse perinatal outcomes by GWG *z* score in normal-weight (*n* = 27,585) (A) and overweight (*n* = 10,989) (B) mothers. Solid lines represent point estimates, and dashed lines represent 95% confidence bands. Outcomes were spontaneous preterm birth (black lines), indicated preterm birth (purple lines), small-for-gestational-age birth (gray lines), large-for-gestational-age birth (gold lines), and unplanned cesarean delivery (pale blue lines). Probabilities were estimated by using log-binomial regression and were set at the population average for maternal race-ethnicity, parity, age, smoking status, marital status, education, and height. GWG, gestational weight gain.

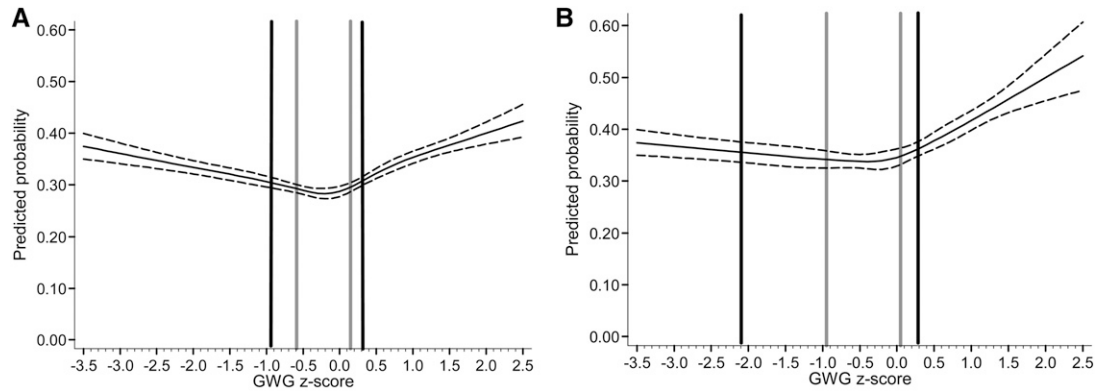


FIGURE 3. Adjusted predicted probability of any one of 5 adverse perinatal outcomes by GWG z score in normal-weight ($n = 27,585$) (A) and overweight ($n = 10,989$) (B) mothers. Solid lines represent point estimates, and dashed lines represent 95% confidence bands. The composite outcome included spontaneous preterm birth, indicated preterm birth, small-for-gestational-age birth, large-for-gestational-age birth, and unplanned cesarean delivery. Probabilities were estimated by using log-binomial regression and were set at the population average for maternal race-ethnicity, parity, age, smoking status, marital status, education, and height. GWG, gestational weight gain.

(adjusted RR: 1.06; 95% CI: 1.04, 1.09). These weight-gain z scores are indicated on the figure by the vertical gray line (Figure 3A). The range between these points (z scores from -0.59 to $+0.15$, which corresponded to 13.2–17.3 kg at 40 wk of gestation) represents the gestational weight gain range in which risks are not meaningfully increased compared with lowest observed risk in the population. In contrast, gestational weight-gain z scores less than -0.59 and greater than $+0.15$ were associated with a degree of risk that was meaningfully higher than risk at the weight gain associated with the best-observed outcomes.

When we repeated this process by using a 20% noninferiority margin, the range of recommended gestational weight gain widened to z scores of -0.97 to $+0.33$ or 11.3–18.4 kg at 40 wk of gestation. This range is indicated by the vertical black lines in Figure 3A.

For overweight women, the relation between gestational weight-gain z scores and individual adverse outcomes that made up the composite is shown in Figure 2B. Directions of associations were similar to those in normal weight women, except that risk of indicated preterm birth in overweight women rose steadily with increasing weight gain at weight-gain z scores of 0 and higher.

When we examined risk of the composite adverse outcome in overweight women (Figure 3B), excess risk associated with low weight gain was less pronounced than for normal weight women, whereas risks associated with high weight gain were more pronounced. In addition, lowest risk of an adverse outcome (ie, 33.7/100), was greater than lowest observed risk in normal-weight women (28.3/100) and occurred at a lower z score (-0.4 compared with -0.2 , respectively). The more-flattened slope of risk curve to the left of the nadir in overweight women resulted in a broader recommended weight-gain range with a 10% noninferiority margin (z scores from -0.95 to $+0.05$, which corresponded to 9.7–16.2 kg at 40 wk of gestation). This 6.5-kg gestational weight-gain range was more than 50% greater than the width of the range in normal-weight women (Table 2).

With a 20% noninferiority margin in overweight women, z-score thresholds were at -2.11 and $+0.29$, which corresponded to a 13.7-kg range from 4.4 to 18.1 kg at 40 wk. As seen in Figure 3B, the broader recommended weight-gain range by using this noninferiority margin was primarily due to a decreased threshold at the lower limit.

TABLE 2

Values for establishing a recommended range of weight gain in pregnancy in 38,574 singleton births at the Magee-Womens Hospital in Pittsburgh, PA, 2003–2010

BMI category	Lowest risk (/100)	Weight-gain z score of lowest risk	Margin of noninferiority	Lower margin of recommended range		Upper margin of recommended range		Corresponding range ¹
				RR (95% CI) ²	Weight-gain z score	RR (95% CI) ²	Weight-gain z score	
Normal weight	28.3	−0.2	10%	1.05 (1.01, 1.09)	−0.59	1.06 (1.04, 1.09)	0.15	13.2–17.3
			20%	1.11 (1.04, 1.19)	−0.97	1.13 (1.08, 1.19)	0.33	11.3–18.4
Overweight	33.7	−0.4	10%	1.02 (0.95, 1.09)	−0.95	1.05 (1.0, 1.09)	0.05	9.7–16.2
			20%	1.08 (0.98, 1.19)	−2.11	1.11 (1.04, 1.19)	0.29	4.4–18.1

¹ Range assumes 40 wk of gestation at delivery.

² RRs were estimated from a log-binomial regression model adjusted for the year of birth, race-ethnicity, parity, marital status, education, insurance status at delivery, smoking, maternal age, and height.

DISCUSSION

In this article, we illustrated how noninferiority margins from clinical trials can be adapted to help establish the recommended range of pregnancy weight gain. Our method offers several advantages over existing approaches, which typically involve post hoc decision making that is based on subjective visual interpretation of graphs. In our approach, the choice of noninferiority margins was made before the data analysis, which made the identification of cutoffs systematic and reproducible. This approach also ensured that the choice of the acceptable degree of increased risk would be consistent when establishing recommendations for different prepregnancy BMI groups.

Optimal weight-gain ranges identified through our analyses should be viewed as an illustration of the methodology and not as substantive recommendations. For simplicity, we chose 10% and 20% noninferiority margins. In practice, noninferiority margins should be chosen on the basis of a combination of expert opinion and statistical reasoning. For this approach to have a maximum impact on weight-gain policy making, researchers with expertise in eliciting health utilities and preferences are needed to work with women, clinicians, and public health professionals in systematically establishing the minimum clinically important difference in risk of various maternal and child health outcomes. Even so, our use of these margins raised several interesting points. First, the 20% margin produced a recommended weight-gain range that was comparable to the current IOM recommendations for normal-weight women (11.3–18.4 compared with 11.5–16 kg, respectively) but was considerably broader for overweight women (4.4–18.1 compared with 7–11.5 kg, respectively) (1). These results suggested that the IOM may have (knowingly or unknowingly) assumed different tolerable degrees of increased risk of normal-weight and overweight women. If 20% is accepted as the threshold at which risk becomes meaningfully increased, the recommended weight-gain ranges for overweight women may be too narrow.

In this study, we used a composite outcome that was based on 5 health outcomes that have been consistently shown to be associated with gestational weight gain. Along with maternal postpartum weight retention, these outcomes were the primary health outcomes considered by the 2009 IOM committee to establish their current recommendations. However, substantive research seeking to establish optimal thresholds would have ideally included a more-comprehensive list of adverse outcomes including long-term maternal and offspring obesity, preeclampsia, gestational diabetes, and infant death. We also used an equally weighted composite outcome to illustrate our proposed approach in the absence of a widely accepted approach for assigning a greater weight to more severe components of the composite outcome (11, 12), but noninferiority margins could equally be applied if outcomes were weighted by using one of several proposed scoring tools (13–15). Although our proposed methodology may help to improve the scientific rigor of the guideline selection process, the methodology is not sufficient by itself, and work to determine how and if different health outcomes should be weighted is an important area for future research.

The range of recommended weight gain obtained through this approach will be influenced by the study's sample size. Smaller studies will tend to have wider CIs that result in a smaller range within which CIs do not exceed the margin of noninferiority.

The practical implication is that ranges of recommended weight gain will be conservative; these ranges will only include values known with reasonable certainty to have risks that are not meaningfully higher than the nadir. We believe that this outcome is an advantage of the approach because public health guidelines should not be based on weight-gain values for which there is uncertainty as to whether risks are meaningfully increased. Our approach helps to make explicit that we need sufficient statistical precision around point estimates to determine with certainty where risks truly become increased.

Because of the impact of the sample size on the recommended range of weight gains, investigators should perform preliminary calculations to estimate the likely degree of statistical precision (width of CIs) in their studies. As with other sample size calculations, these calculations would be based on pilot data or previous studies. A simple approach would be to estimate the number of women and number of adverse outcomes in gestational weight-gain categories in which risks are hypothesized to become meaningfully increased and calculate the point estimate and width of its expected CIs. More sophisticated estimates could be obtained by using simulation. The minimum width of CIs needed for a study could be informed by, eg, increments in which weight-gain guidelines are made (currently, recommendations are rounded to the nearest 0.5 kg), and thus, greater precision than this would not be necessary because it would not lead to different recommendations.

We recognize that women, clinicians, and public health professionals may differ in their opinions on the minimum difference in risk between groups they consider to be clinically important (ie, individuals may select different noninferiority margins). As a result, we recommend that researchers who seek to inform pregnancy weight-gain guidelines present their study results in a manner that allows ranges that reflect different noninferiority margins to be identified. For example, the presentation of RRs or risk differences associated with many different weight-gain values under "Supplemental data" in the online issue, not just those for which the 95% CI first crosses the 10% and 20% margins, allows readers to establish ranges on the basis of alternative margins.

The methodology described in this article should be viewed as a starting point for future testing and potential modification. We used a single data set with 5 outcomes for illustrative purposes, but additional studies should be used to establish definitive public health guidelines. Guidelines should be informed by studies that include a broad range of short- and long-term health outcomes that control for relevant confounders and are based on high-quality data to avoid biased estimates of risk that are the result of measurement error. Studies should be based on diverse, generalizable cohorts with a broad range of gestational weight gain and prepregnancy BMI. Work to establish methodologies that combine multiple different cohorts through meta-analysis to create such a cohort would be valuable.

In conclusion, noninferiority margins used in clinical trials have a number of potential applications in public health. In this study, we illustrated how they can be used to provide a systematic, reproducible approach for establishing thresholds for recommended gestational weight-gain ranges. In addition, the formal process of eliciting noninferiority margins may help to engage women, policy makers, and clinicians early in the research process and help to ensure that study results are presented

in a way that is most informative for policy-making. This approach is not intended to replace expert opinion but to provide quantitative evidence to better inform expert opinion. Expert evaluation of the study quality that underlies the ranges (eg, the inclusion of all clinically important outcomes and generalizability of study population) will always be needed. Other public health uses could potentially include establishing optimal ranges of nutrient intake for public health dietary guidelines or determining ranges of healthy weights in children and adults. The use of this method will help to ensure that recommended ranges in public health guidelines are selected in an evidence-based manner and directly reflect the range of values associated with acceptably low increases in risks as predetermined by clinicians and public health officials.

We thank Barbara Abrams (University of California, Berkeley) and Kathleen Rasmussen (Cornell University) for their insightful comments on earlier versions of the manuscript.

The authors' responsibilities were as follows—LMB: provided the study data set and performed statistical analyses; and both authors: jointly designed the research, wrote the manuscript, took responsibility for the final content of the manuscript, and read and approved the final manuscript. Neither author declared a conflict of interest.

REFERENCES

1. Weight Gain During Pregnancy IOM. Reexamining the guidelines. Washington, DC: National Academies Press, 2009.
2. Jones B, Jarvis P, Lewis JA, Ebbutt AF. Trials to assess equivalence: the importance of rigorous methods. *BMJ* 1996;313:36–9.
3. Villar J, Ba'aqeel H, Piaggio G, Lumbiganon P, Belizan JM, Farnot U, Al-Mazrou Y, Carroli G, Pinol A, Donner A, et al. WHO antenatal care randomised trial for the evaluation of a new model of routine antenatal care. *Lancet* 2001;357:1551–64.
4. Julious SA, Owen RJ. A comparison of methods for sample size estimation for non-inferiority studies with binary outcomes. *Stat Methods Med Res* 2011;20:595–612.
5. Piaggio G, Elbourne DR, Altman DG, Pocock SJ, Evans SJ. Reporting of noninferiority and equivalence randomized trials: an extension of the CONSORT statement. *JAMA* 2006;295:1152–60.
6. Ross S, Milne J, Dwinell S, Tang S, Wood S. Is it possible to estimate the minimal clinically important treatment effect needed to change practice in preterm birth prevention? Results of an obstetrician survey used to support the design of a trial. *BMC Med Res Methodol* 2012;12:31.
7. Harrell FE Jr, Lee KL, Pollock BG. Regression models in clinical studies: determining relationships between predictors and response. *J Natl Cancer Inst* 1988;80:1198–202.
8. 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. *Am J Clin Nutr* 2013;97:1062–7.
9. Hadlock FP, Harrist RB, Martinez-Poyer J. In utero analysis of fetal growth: a sonographic weight standard. *Radiology* 1991;181:129–33.
10. World Health Organization. Physical status: the use and interpretation of anthropometry. Report of a WHO Expert Committee. *World Health Organ Tech Rep Ser* 1995;854:1–452.
11. Hoeldtke NJ. Dubious validity of novel maternal-fetal outcome model. *Am J Obstet Gynecol* 2001;185:256–7.
12. Yancey MK, Nielsen PE. Disappointment with maternal-fetal outcome model. *Am J Obstet Gynecol* 2001;185:257; author reply 258.
13. Novicoff WM, Wagner DP, Knaus WA, Kane EK, Cecere F, Draper E, Ferguson JE 2nd. Initial development of a system-wide maternal-fetal outcomes assessment program. *Am J Obstet Gynecol* 2000;183:291–300.
14. Verma A, Weir A, Drummond J, Mitchell BF. Performance profile of an outcome measure: morbidity assessment index for newborns. *J Epidemiol Community Health* 2005;59:420–6.
15. Oken E, Kleinman KP, Belfort MB, Hammitt JK, Gillman MW. Associations of gestational weight gain with short- and longer-term maternal and child health outcomes. *Am J Epidemiol* 2009;170:173–80.