Birthweight charts in the study fetal growth: current limitations and potential alternatives

Jennifer Anne Hutcheon

A thesis submitted to the Faculty of Graduate and Postdoctoral Studies in partial fulfillment of the requirements for the degree of Doctor of Philosophy (Ph.D.)

Department of Epidemiology, Biostatistics, and Occupational Health
McGill University
Montreal, Canada
February 2009

Abstract/Résumé

Birthweight-for-gestational-age charts have long been used to identify infants at increased risk of adverse perinatal outcomes due to fetal growth restriction. Despite their widespread use, conventional birthweight charts have several important limitations. Although the pathological process of interest is the longitudinal process of poor fetal *growth*, birthweight charts classify infants based only on the cross-sectional measure of *weight*. As a result, infants that are small, but healthy, can be inappropriately identified as being at increased risk of adverse outcomes. Further, most conventional birthweight charts are created from the weights of livebirths at each completed week of gestation. Since the weights of livebirths at preterm ages are known to be smaller than the weights of ongoing pregnancies of similar gestational age, conventional birthweight charts are not representative of the total cohort at early gestational ages. Given these known limitations, the goals of this thesis were 1) to evaluate the potential for bias arising from the use of conventional birthweight-for-gestational-age percentiles in epidemiologic studies of fetal growth restriction and, 2) to evaluate two alternatives to conventional birthweight-for-gestational-age percentiles: "customized" birthweight percentiles and "conditional" fetal growth percentiles.

This thesis first outlines the theoretical bias created by the classification of "small-for-gestational-age" (SGA, a weight below the 10^{th} percentile for gestational age) of conventional birthweight charts at preterm gestational ages. Using simulations, the impact of this theoretical bias on studies of risk factors for fetal growth restriction is quantified and shown to be of sufficient magnitude to impact substantive conclusions. Next, the use of "customized" percentiles, birthweight percentiles that have been adjusted to account for maternal influences on fetal growth such as height, parity, ethnicity or pre-pregnancy body mass index, are examined. Although customized percentiles are widely believed to be superior to conventional birthweight-for-gestational-age percentiles in identifying infants at increased risk of adverse perinatal outcomes, we demonstrate that their apparent benefits are primarily due to their use of an intrauterine standard at preterm ages and that the process of adjusting for maternal characteristics added very little to the identification of perinatal risk. Finally, "conditional" fetal growth percentiles, percentiles calculated given (conditional on) an infant's weight earlier in pregnancy, are assessed. Despite having been proposed in the statistical literature as a more

appropriate measure of fetal growth restriction, we were unable to detect an improvement over conventional percentiles in identifying adverse perinatal outcomes related to fetal growth restriction.

Les graphiques de poid*s* à la naissance pour l'âge gestationnel ont longtemps été utilisés afin d'identifier les nouveau nés à risque élevé de mortalitée et à morbidité périnatale en raison d'hypotrophie foetale. Malgré leur utilisation étendue, les graphiques de poid*s* à la naissance pour l'âge gestationnel conventionnel ont plusieurs restrictions importantes. Bien que le processus pathologique d'intérêt soit le processus longitudinal de la pauvre croissance foetale, les graphiques de poid***s*** à la naissance classifient les nouveau-nés seulement par la mesure de poids. Conséquemment, les nouveau-nés qui sont petits, mais en bonne santé, peuvent être inopportunément identifiés comme étant à risque augmenté de mortalité/morbidité. De plus, les graphiques de poid***s*** à la naissance conventionnels sont créés avec seulement les poids des nouveau-nés à chaque semaine complète de gestation. Puisqu'il est connu que le poids d'un nouveau-né à l'âge préterme est plus petit que le poids de grossesse en cours d'âge gestationnel semblable, les graphiques de poid***s*** à la naissance pour l'âge gestationnel conventionnel ne sont pas représentatifs de la cohorte totale à l'âges gestationnel préterme. Étant donné ces restrictions connues, les objectifs de cette thèse étaient 1) évaluer le potentiel pour la partialité avec l'utilisation de centiles de poid***s*** à la naissance conventionnels dans les études épidémiologique d'hypotrophe foetale et, 2) évaluer deux alternatives aux graphiques de poid***s*** à la naissance conventionnels : les centiles de poid***s*** à la naissance "personnalisés" et les centiles de croissance foetaux "conditionnels".

Cette thèse débute en décrivant la partialité théorique créée par la classification de "Poids à la naissance faible" (un poids au-dessous du 10ème centile pour l'âge gestationnel) des graphiques de poid***s*** à la naissance pour *l'*âge gestationnel préterme. En utilisant des simulations, l'impact de cette partialité théorique sur les études de facteurs de risque pour l'hypotrophe foetale est quantifié et est suffisante pour avoir un impact sur les conclusions substantives. Ensuite, l'utilisation de centiles " personnalisés " (les centiles de poid***s*** à la naissance qui ont été réglés pour représenter des influences maternelles sur la croissance foetale comme la grandeur, la parité, l'ethnicité ou l'index de masse corporelle pré-grossesse) est examinée. Bien que l'on croit largement que les centiles " personnalisés " sont supérieurs aux centiles de poid***s*** à la naissance conventionnels dans l'identification des nouveau-nés à risque augmenté de mortalité/morbidité périnatale, nous démontrons que leurs avantages apparents sont essentiellement en raison de leur utilisation d'une norme intrautérine à l'âge préterme et que le processus de réglage pour les caractéristiques maternelles a ajouté très peu à l'identification de risque périnatal. Finalement,

les centiles de croissance foetaux "conditionnels", les centiles calculés donné (conditionnel sur) le poids d'un nouveau né plus tôt dans la grossesse, sont évalués. Malgré avoir été proposé dans la littérature statistique comme une mesure plus appropriée de l'hypotrophie foetale, nous étions incapables de découvrir une amélioration sur les centiles conventionnels dans l'identification des nouveau-nés à risque augmenté de mortalité/morbidité périnatale.

Acknowledgements

This thesis would not have been possible without the guidance and support of many individuals. First and foremost, I thank my thesis supervisor, Robert Platt. He has always made his students' education a priority in his schedule, and I have learned as much from our many informal conversations and email exchanges over the past four years as I have from any more formal teaching. I appreciate that he gave me the flexibility to take this thesis in my own direction, while still making sure I didn't veer off course. I could not have asked for a better advisor throughout my doctoral training. I also thank my co-supervisor, Grace Egeland. In addition to her epidemiologic input, what I appreciate about working with her is the creativity and big picture thinking that she brings to all her projects.

I am very grateful to my committee members, Sara Meltzer and Lucie Morin, for taking time out of their busy schedules to agree to be on my committee. Their clinical insights on my doctoral research were invaluable. My co-authors, Xun Zhang, Michael Kramer, and Sven Cnattingius were equally helpful with their advice, and collaborating with them was a real pleasure.

I feel lucky to have spent the past few years working with Xun, Mourad, Helen, Adriana, Rob, and Mike in the perinatal Epidemiology group at McGill. The support, collaboration, exchange of ideas, and humour in the group create an environment that I will miss very much. I will also miss the faculty and my fellow students in Epidemiology & Biostatistics at McGill. I would like to thank in particular Jim Hanley and Rebecca Fuhrer for their mentorship.

I thank the wonderful city that is Montreal for taking me in for a few years. Its energy and life undoubtedly helped keep me sane throughout the ups and downs of doctoral research. *Les filles*, I don't know what I'm going to do with myself on Thursday nights.

I thank Jason for his love, perspective, and countless hours spent waiting for late VIA trains post-call. Finally, I thank with love my parents, my sister, and my Nana for their endless support and for instilling in me an appreciation for learning.

Contribution of authors

Manuscript 1.

Hutcheon JA and Platt RW. The missing data problem in birth weight percentiles and thresholds for "small-for-gestational-age". Am J Epidemiol 2008; 167:786-792.

The material in this methodological manuscript is based on a series of discussions and emails between Dr. Platt and myself, which I consolidated into manuscript form. I wrote the first draft of the manuscript, and produced all graphs and calculations. The manuscript evolved through numerous revisions, in which Dr. Platt was central in suggesting material to remove or ideas to develop further. He had an important role in editing the final version of the manuscript both for language as well as content.

Manuscript 2.

Hutcheon JA, Zhang X, Cnattingius S, Kramer MS, and Platt RW. Customized birth weight percentiles: does adjusting for maternal characteristics matter? BJOG 2008; 115:1397–1404.

This manuscript was an extension of work previously done by my co-authors.[1] The study dataset, provided by Dr. Cnattingius, had been prepared for earlier analyses by Dr. Zhang. I developed the research objectives and study protocol, performed all analyses presented in this paper, and wrote the first draft of the manuscript. Drs. Zhang, Cnattingius, Kramer, and Platt reviewed the protocol, were actively involved in debate over the interpretation of results, and edited all versions of the manuscript.

Manuscript 3

Hutcheon JA, Egeland GM, Morin L, Meltzer SJ, Jacobsen G, Platt RW. The predictive ability of conditional fetal growth percentiles. To be submitted to Paediatric and Perinatal Epidemiology.

The research objectives and protocol for this manuscript were conceived by myself. The databases used in this study were provided to me by Dr. Lucie Morin and Mme Danielle Vallerand of the Royal Victoria Hospital. I merged the databases required to produce the study cohort, performed all analyses, and wrote the first draft of the paper. Drs. Morin and Meltzer

provided invaluable clinical expertise on the obstetrical ultrasound procedures used at the Royal Victoria Hospital and the clinical implications fetal growth restriction. Drs. Egeland, Jacobsen, and Platt provided advice on the study design and methodologies used in the study. Dr. Platt additional provided statistical expertise on the calculations required to produce conditional percentiles.

Statement of originality

The work presented in this thesis represents an original contribution to the field of perinatal epidemiology. To my knowledge, the potential for bias arising from the use of "small-for-gestational-age" as an outcome in epidemiologic studies of fetal growth restriction has not previously been considered. Although the properties of customized birthweight percentiles had previously been examined by my co-authors, the study in this thesis was the first to explicitly compare the ability of customized birthweight standards in predicting risk of perinatal mortality to that of an ultrasound estimated- fetal-weight-for-gestational age chart as well as a conventional birthweight chart. Finally, our evaluation of conditional fetal growth percentiles was the first study of which I am aware to explore the ability of conditional fetal growth percentiles to predict serious perinatal morbidity and mortality related to fetal growth restriction.

While I have received guidance from my committee members and co-authors on statistical, clinical and methodological aspects of this thesis, I declare that the conception, execution, and drafting of the work in this thesis were my own.

Statement of financial support

This thesis was funded by a Doctoral Research Award from the Canadian Institutes of Health Research (Clinical Research Initiative), for which I am very grateful. The Research Institute of the Montreal Children's Hospital provided travel awards for conference presentations of portions of this work.

Table of Contents

1. Introduction

Birthweight for gestational age has long been recognized to be one of the most powerful predictors of perinatal outcome.[2] Infants born smaller than their peers of similar gestational age are at increased risk of perinatal mortality and other adverse outcomes such as neonatal encephalopathy and multi-organ dysfunction.[3] These sequelae of intrauterine growth restriction are believed to result from an inadequate transfer of oxygen and nutrients across the placenta to the fetus, leading to hypoxia and malnutrition.[3] The risks of adverse perinatal outcomes increase exponentially with decreasing birthweight,[4] re-enforcing the importance of promoting optimal fetal weight gain during pregnancy.

Early classification schemes to assess the risks associated with fetal growth restriction were based solely on birthweight. The 1950 World Health Organization Expert group on Prematurity, for example, endorsed the use of the classification "Low birth weight" (LBW), defined as "a birthweight below 2500 grams".[5] Such classifications, however, failed to distinguish between infants who were small due to insufficient time in utero and infants who were small because of a poor rate of growth. [6] Birthweight-for-gestational- age percentiles were developed to better separate the risks of prematurity from the risks of poor intrauterine growth.[7, 8] In these charts, the weights of all births at each week of gestation are converted into percentiles, and those below a certain threshold of the population, typically the 10th percentile, are classified as "small-for-gestational-age" (SGA) and considered to be at increased risk of adverse perinatal outcome due to poor fetal growth. Birthweight-for-gestational-age charts have been created for numerous populations throughout the world,[9-14] and at present, are the cornerstone of screening for infants with poor fetal growth.

While birthweight-for-gestational-age charts have been widely adopted into clinical care and public health practice, they are not without their limitations. Birthweight charts are intended to identify infants with poor fetal *growth*, but what they actually identify is small fetal *size*.[15] The use of the cross-sectional measurement of size instead of the longitudinal measurement of growth means that birthweight-for-gestational-age charts are unable to distinguish between infants who are small but growing well and infants who have failed to meet their own growth potential. The classification of "small-for-gestational-age" will include many infants who are

small-but-healthy, and will fail to include infants whose weight is above the population 10th percentile but whose growth has been restricted *in utero*. Further, the percentiles from conventional birthweight-for-gestational-age charts are based on the distribution of *birthweights* at a given gestational age, not the weight distribution of the total cohort of ongoing pregnancies. Since preterm infants are known to be smaller than their *in utero* peers,[16-19] the weight percentiles of births at early gestational ages are not representative of the weight percentiles of the total population. As a result, any preterm infants whose weight was in the smallest 10 percent of pregnancies of similar gestational duration, but not the smallest percent of preterm births, will not be classified as "small-for-gestational-age".

When used to identify infants at increased risk of adverse outcomes in clinical practice, the limitations of conventional birthweight-for-gestational-age charts are not unrecognized. When used to study the etiology of fetal growth restriction at the population level, however, the limitations of conventional birthweight charts do not appear to have been considered. "Small-for-gestational-age" is perhaps one of the most commonly studied outcomes in perinatal epidemiology, used to answer questions such as "To what extent does second hand smoke restrict fetal growth?" or "Is genotype X associated with fetal growth restriction?" Given the known limitations of conventional birthweight-for-gestational-age charts, how valid is the use of "small-for-gestational-age" as a outcome to study risk factors for fetal growth restriction? What impact, if any, could it have on our understanding of the causes of poor fetal growth? The first objective of this thesis was to evaluate the potential for bias arising from the use of birthweight-for-gestational-age charts (and in particular, the resulting classification of "small-for-gestational-age") in epidemiologic studies of fetal growth restriction. The first manuscript of the thesis, "The missing data problem in birth weight percentiles and thresholds for "small-for-gestational-age"' (*Am J Epidemiol* 2008; 167:786-792), outlines the methodological concerns that arise from using conventional birthweight-for-gestational-age charts in studies of fetal growth restriction, and uses a simple simulation to illustrate the potential impact these concerns may have on substantive conclusions.

Conventional birthweight-for-gestational charts may suffer from limitations both for clinical and research uses, but are there any better alternatives? The second objective of this thesis was to explore alternatives to the classification of "small-for-gestational-age" obtained from

conventional charts. Evaluations of two existing alternatives, so-called “customized” birth weight percentiles and “conditional” growth percentiles, are presented in the second and third manuscripts of the thesis, entitled “Customised birthweight percentiles: does adjusting for maternal characteristics matter?” (*BJOG* 2008; 115:1397–1404) and “The predictive ability of conditional fetal growth percentiles” (to be submitted to *Paediatr Perinatal Epidemiol*), respectively. Finally, after having considered both the limitations of existing birthweight-for-gestational-age charts as well as several alternatives, the thesis will conclude with a consideration of the most appropriate classification of fetal growth restriction in view of the results presented, as well as recommendations for future work in the field.

2. Literature Review

2.1 Intrauterine growth restriction

2.1.1 Etiology

Intrauterine growth restriction (IUGR) is a term used to describe the manifestation of fetal, placental, or maternal disorders that interfere with normal fetal growth. As such, it is not a disease *per se*, but rather a phenotype, with multiple causes and multiple clinical sequelae. Although chromosomal abnormalities and congenital anomalies are well-known causes of suboptimal fetal growth,[20, 21] in common practice the term is generally used to describe a restriction in fetal growth resulting from inadequate transfer of oxygen and/or nutrients across the placenta to the fetus.[22]

Fetal growth restriction arising from maternal causes include those due to decreased utero-placental blood flow, decreased oxygen carrying capacity, and suboptimal maternal nutritional status.[23] Decreased utero-placental blood flow, leading to dysfunctional oxygen and nutrient delivery, is commonly caused by maternal vascular diseases including hypertension in pregnancy or pre-eclampsia, chronic hypertension, chronic renal disease, autoimmune disorders, or diabetic vasculopathy.[24] Factors resulting in reduced maternal oxygenation include high altitude or hypoxic disorders such as hemoglobinopathies and anemias.[24] Maternal smoking is believed to cause IUGR both by reducing uterine blood flow and impairing fetal oxygenation.[23] Maternal malnutrition, as seen through a low pre-pregnancy weight or inadequate weight gain during pregnancy, results in growth restriction due to lack of nutrient availability to the fetus, though the extremes of maternal malnutrition needed to have a meaningful impact on fetal growth would rarely be encountered in developed countries.[21] Finally, use of drugs such as heroin, cocaine, and anti-convulsants have also been associated with fetal growth restriction.[21]

Fetal etiologies of IUGR include multiple pregnancy and infections.[3, 23] Infections of concern in the developed world include toxoplasmosis and cytomegalovirus, while malaria is a common cause of intrauterine growth restriction in many developing countries.[23] Direct cytolysis and localized necrosis in the fetus is believed to be the mechanism of decreased fetal growth due viral infections such as cytomegalovirus, while protozoan infections such as malaria and toxoplasmosis cause fetal growth restriction through pathological damage to the placenta,

disrupting blood flow and nutrient/oxygen transfer. Multiple pregnancies are often associated with fetal growth restriction, with up to 15-30% of twins being growth restricted, likely due to decreased substrate.[21, 25]

Abnormal development of the placenta affects roughly one third of pregnancies with IUGR.[22] A small placental mass (in absolute size, or relative to fetal size) limits the transfer and availability of susbtrates to the fetus, restricting fetal growth.[26] Additional placental or cord abnormalities such as chorioangiomata (tumour), infarction, placenta previa, and circumvallate placentae are also thought to compromise nutrient and oxygen transport to the fetus.[23, 27]

2.1.2 Symmetrical versus asymmetrical intrauterine growth restriction

The various etiologies of intrauterine growth restriction are believed to result in different patterns of fetal growth restriction.[3, 21, 22] Symmetric fetal growth restriction is characterized by smaller sizes of all body components, including weight, head circumference, and length. This form of growth restriction is thought to result from an onset early in gestation, during a period when the fetus is growing primarily through cell division. Etiologies believed to result in symmetrical growth restriction include infection, drugs, and congenital malformations.[3, 21, 22] Asymmetrical growth restriction is thought to result from a lack of metabolic substrates during the second half of pregnancy when fat deposition normally occurs. Growth restriction during this period is believed to be exhibited through a sparing of the head circumference and length, but decreased abdominal circumference (due to inadequate deposition of subcutaneous fat and decreased liver size). Asymmetrical growth restriction is generally attributed to etiologies related to decreased uteroplacental perfusion, such as maternal vascular disease or placental abnormalities.[3, 21, 22]

Although the idea of distinguishing between symmetric and asymmetric fetal growth restriction is widespread,[3, 21, 22] existing evidence does not support the importance of this distinction. Longitudinal ultrasound studies have found no difference in the timing of onset of fetal growth restriction between asymmetrically and symmetrically small infants,[28] and perinatal outcomes are similar between symmetrically and asymmetrically small infants once severity of growth restriction is taken into account.[29] Asymmetry is likely simply a proxy for severe growth restriction, and appears to have little clinical importance by itself.

2.1.3 Clinical manifestations of fetal growth restriction

Irrespective of the etiology, the inadequate transfer of oxygen and nutrients across the placenta characteristic of fetal growth restriction leads to decreased glucose availability, the fetus' primary energy substrate.[30] Inadequate fetal glucose availability stores triggers a switch to alternative energy sources of protein and fat.[31, 32] Endogenous muscle is broken down for amino acids, and diversion of fatty acids as energy sources results in low fat deposition and decreased rate of fetal weight gain.[22, 32] Ketone bodies, a by-product of fatty acid metabolism, accumulate in the blood and lead to the development of metabolic acidosis.[31] The fetus' limited oxygen supply further limits its capacity for oxidative metabolism, leading to increased production of lactic acid and worsening metabolic acidosis.[33] The severe clinical sequelae of perinatal hypoxia and acidosis include neonatal encephalopathy (seizures, respiratory difficulties, hypotonia, and decreased level of consciousness), renal failure due to acute tubular necrosis, pulmonary vasoconstriction (with risk of persistent pulmonary hypertension), decreased cardiac contractility, and ultimately, intrauterine fetal demise.[3, 34] Additional metabolic complications of the growth restricted infant include hypothermia, polycycthemia, neonatal hypoglycaemia, and hypocalcemia.[34]

2.2 Birthweight-for-gestational-age charts

2.2.1 Birthweight charts in the identification of intrauterine growth restriction

The strong association between small fetal size, the endpoint of suboptimal intrauterine development, and perinatal mortality[2, 35, 36] has led to the widespread use of low birthweight as a marker of fetal growth restriction.[2, 36] The importance of birthweight as a predictor of perinatal outcome is illustrated in Figure 2-1, where risks of risks of neonatal mortality among births in the United States are observed to increase exponentially with decreasing birthweights.

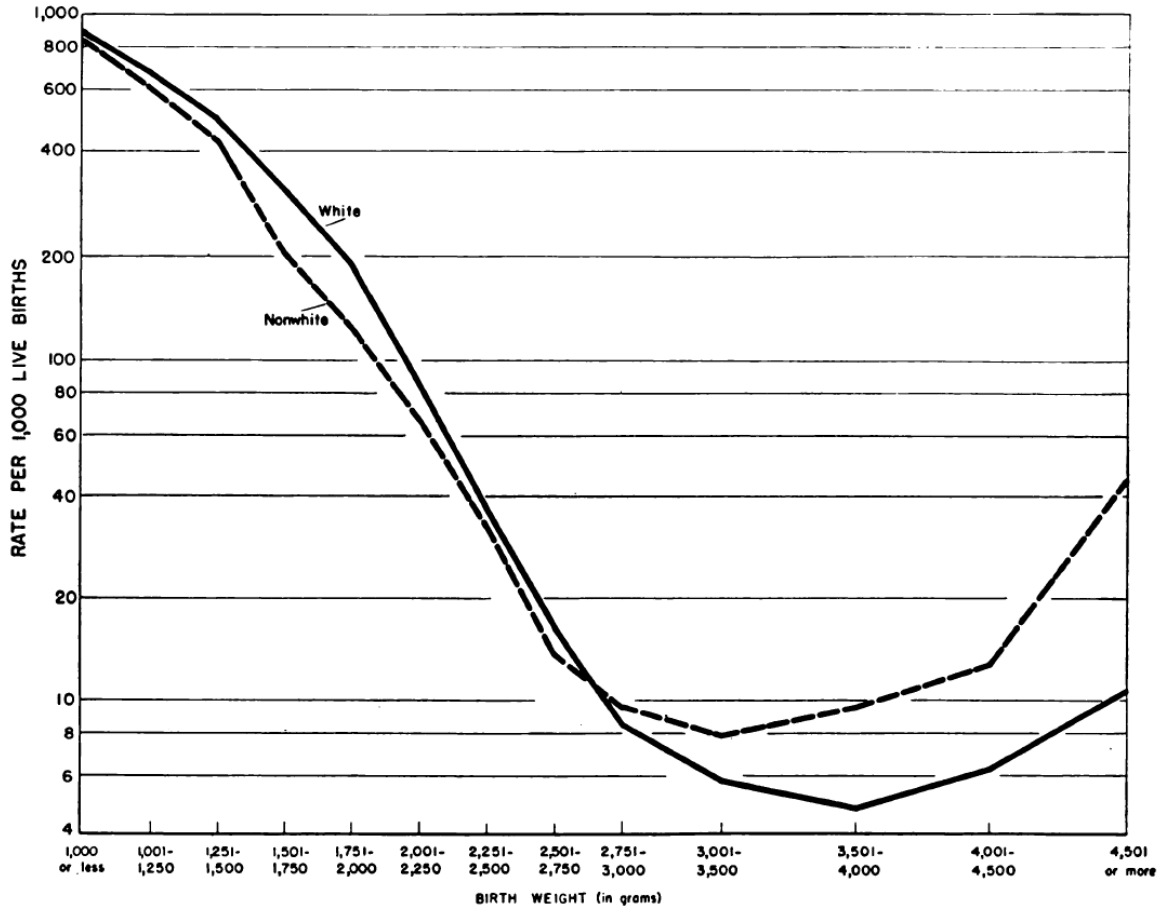


Figure 2-1. Neonatal mortality rates among single births in hospitals. By detailed birth weight and race, United States, January 1 to March 31 1950.
Reprinted from Pediatrics, "Birth weight and survival of the newborn" 1954;14:396 (p.399), with permission from the American Academy of Pediatrics.

Early birthweight classifications to identify high-risk infants were based solely on absolute weight, irrespective of gestational age of the infant. The 1950 World Health Organization Expert Group on Prematurity,[5] for example, endorsed a classification scheme in which infants weighing less than 2500grams at birth (Low Birth Weight, LBW) were considered "premature". Such classifications, however, failed to distinguish between infants born small because they were born too soon (preterm) and infants born small because of restricted intrauterine growth.[2, 35, 37] As described in a 1967 editorial in the Journal of Pediatrics,[6] "The need for a new and appropriate classification appears to be generally accepted. The terms, prematurity and premature infant, have been useful for a long time in identifying the infant of low birth weight whose risks of morbidity and nonsurvival are considerably greater than those infants whose birth weights are within the range of 'normal'. As it became increasingly clear that not all low-birth-weight infants were premature by gestational age and that time in utero as well as weight

at birth affect the infant's ultimate prognosis, schematic methods for classification and terminology were proposed which might be more discriminative than the term, premature infant" (p.309).

That 1967 issue of the Journal of Pediatrics contained two such proposed classifications. Yerushalmy [38] proposed a 5-group scheme based on a cross-classification of birthweight and gestational age at birth. Group I consisted of infants of all gestational ages weighing less than 1500 grams. Group II consisted of infants born younger than 37 weeks weighing between 1501-2500 grams, while group III consisted of infants 37 weeks or older weighing between 1501-2500 grams. Group IV consisted of infants born younger than 37 week weighing above 2501 grams, and group V consisted of infants 37 weeks or older weighing above 2501 grams. His choice of categories was supported by calculations of neonatal mortality within each group from a 3 year period in New York, with group I having a neonatal mortality risk of 707.3 per thousand, group II of 104.7 per thousand, group III of 32.0 per thousand, group IV of13.7 per thousand, and group V of 4.7 per thousand.

The second classification, published by Battaglia and Lubchenco,[7] was based on the weight-for-gestational-age percentile chart published 4 years earlier by Lubchenco[8] for a population of births in Denver, Colorado. Infants below the 10th percentile of weight for gestational age were classified as "small-for-gestational-age" (SGA), infants between the 10th and 90th percentiles classified as "appropriate-for-gestational-age" (AGA), and infants above the 90th percentile classified as "large-for-gestational-age" LGA. These 3 weight-for-gestational-age categories were then combined with 3 classifications of gestational age (preterm (defined as infants less than 38 completed weeks), term (38 completed weeks to 42 weeks), and post-term (42 completed weeks or greater)) to produce 9 separate risk categories.

The classification of infants into 9 distinct risk categories proposed by Battaglia and Lubchenco [7] was not widely adopted by the clinical and research communities, but their proposal to classify fetal weight based on "small-for-gestational-age", "appropriate-for-gestational-age", and "large-for-gestational-age" status has since become a standard in the assessment of fetal growth. Birthweight-for-gestational-age charts have been published for numerous countries worldwide[9-14] with a variety of modifications including corrections for error in gestational age[12] or population-specific characteristics such as ethnicity[14, 39] and parity.[11, 14] The sex-specific

birthweight-for-gestational-age chart adopted by Health Canada to assess the weight of Canadian births[12] is shown in Figure 2-2.

The optimal threshold for "small-for-gestational-age" has been investigated by examining risks of perinatal morbidity and mortality among different weight percentile groups. Although researchers have advocated the use of different percentiles such as the 3rd or 15th as the optimal threshold for SGA,[40, 41] the original choice of the 10th percentile remains generally accepted in North America.[21] It should be noted that in Scandinavian countries, however, the use of - 2 standard deviations is commonly used as the threshold for SGA.[16, 42] Alternatively, a reference based on relative risks has also been proposed,[43] where rather than using weight percentiles as thresholds, the weights associated with a 2-fold, 2.5- fold and 3-fold neonatal death risk at each gestational age are used to define SGA. Recently, an outcome-based standard has been created that classifies optimal birthweight based on the risk of serious neonatal morbidity or mortality across different birthweights.[44] The range in which adverse outcomes are lowest (i.e. the range at the bottom of the inverted 'J' relationship between birthweight and risk of adverse outcome (shown in Figure 2-1) in which the slope of risk appears "flattened") was identified within different weeks of gestation, fetal sex and plurality.

2.2.2 Birthweight charts in the study of intrauterine growth restriction

In addition to use in clinical practice to identify individual infants at increased risk of adverse perinatal outcomes, the classification of "small-for-gestational-age" is also widely used by researchers studying the causes and risk factors for intrauterine growth restriction. "Small-for-gestational-age", along with preterm birth, is arguably one of the most commonly studied classifications in perinatal epidemiology. A Medline search of the years 1950 to September Week 2 2008 using the MeSH term "Infant, Small For Gestational Age", for example, revealed nearly 3800 studies using this definition (4567 studies were retrieved with the MeSH term "Premature Birth" and the keyword "Preterm birth"). These range from case-control studies examining genetic risk factors for intrauterine growth restriction (with a case definition established as "Birthweight below the 10th percentile")[45] to cohort studies examining the risk of SGA among women exposed/unexposed to risk factors such as asthma, second-hand smoking, or hypertension.[46-48] In recent years it has also become used in studies of the long-term health effects of antenatal development on risk of chronic disease such as type 2 diabetes and cardiovascular disease.[49-52]

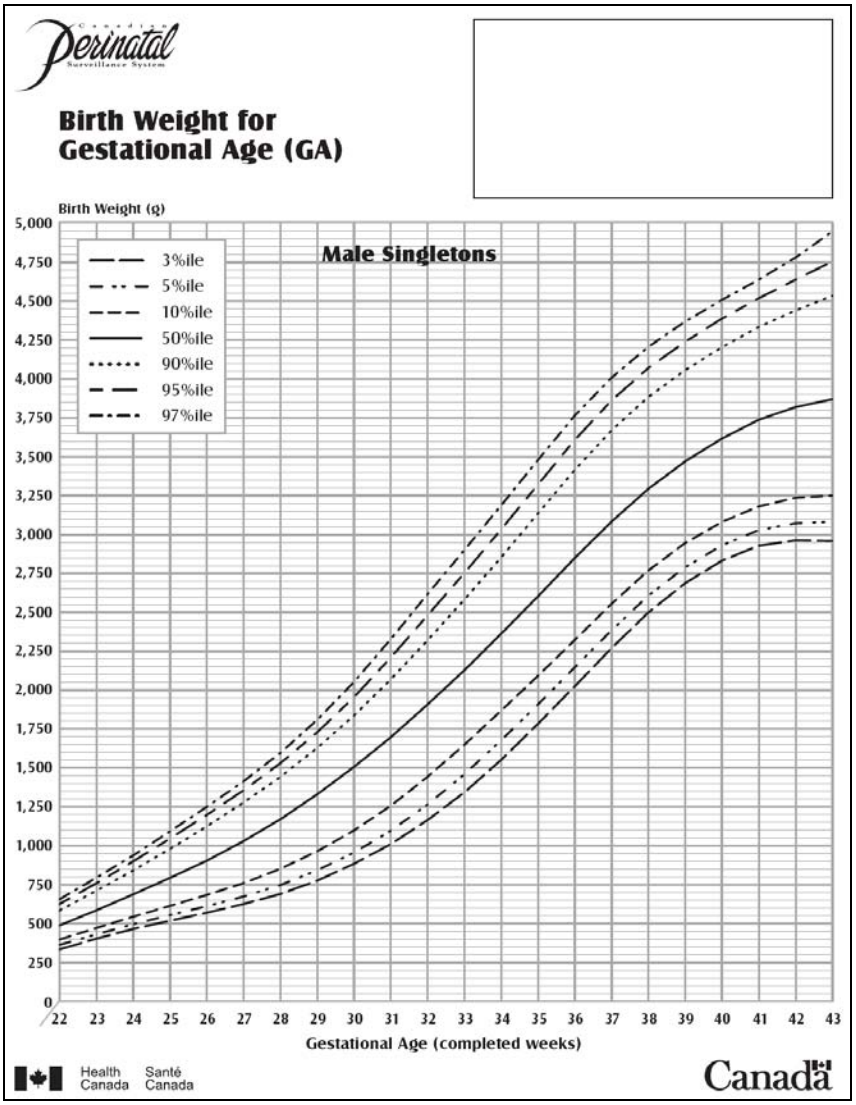


Figure 2-2. Birthweight-for-gestational-age chart for Male Singletons in Canada. Available from the Public Health Agency of Canada website.[53]

2.3 Limitations of birth weight-for-gestational-age charts

2.3.1 Errors in estimation of gestational age

Birthweight-for-gestational-age charts are not without limitations. In order to calculate weight percentiles for each completed week of gestation, accurate information on gestational age is required. Before the widespread availability of obstetrical ultrasound, estimates of gestational age were obtained by calculating the interval between the first day of the last normal menstrual period (LMP) and the day of delivery (under the assumption that women's menstrual cycles are 28 days, and the duration of human pregnancies is 280 days).[22] Because of variability in the

length of women's menstrual cycles, early pregnancy blood spotting, and uncertainty in maternal recall of the dates of their last period, this approach is known to be highly error-prone.[54] When birthweight percentile charts are then created from LMP-dated pregnancies, the error in gestational ages means that the weight distribution within each week of gestation will be artificially wide, or even bimodal.[55] Any artificial widening of the percentile range would be expected to lower the weight of the 10th percentile SGA threshold, which in turn would lead to infants that were truly in the smallest 10 percent of their peers not being identified as SGA. Although recent birthweight charts created from LMP-dated pregnancies[9, 13] (most notably, for the United States population[9]) have attempted to remove implausible estimates of gestational age (by removing observation > 2.5 standard deviations, or by removing infants in the secondary peak of a bimodal distribution), a certain degree of error likely remains in their reference values. This difficulty in obtaining accurate estimates of gestational age is an important reason why early classifications of fetal growth may have been based only on birthweight (i.e. Low birthweight (LBW) defined as a birthweight <2500g). Although they were unable to distinguish infants born too small from infants born too young, their key advantage was that they relied solely on the (accurate) measurement of birthweight, and did not rely on the error-prone measurement of gestational age.

Estimation of gestational age was improved with the introduction of obstetrical ultrasound, which uses fetal biometric measurements in the first half of pregnancy to establish gestational age (usually the crown-rump length at 8-13 weeks, or biparietal diameter from roughly 14-22 weeks).[22, 56, 57] The gold standard for estimation of gestational age is currently either based on ultrasound alone, or based on last menstrual period dating corrected with early ultrasound (i.e. use of LMP unless the discrepancy between the LMP and ultrasound estimates is greater than a pre-specified number of days, in which case the latter estimate is used[58]).[22] Birthweight references for numerous countries,[10, 11 12] (including Canada) are based on pregnancies predominantly dated by ultrasound, or LMP confirmed with early ultrasound. Though an improvement over gestational ages estimated solely from LMP,[59] the use of fetal biometrics measurements to date pregnancies does rely on the assumptions that fetal biometry is strongly correlated with gestational age and that there is minimal variability in fetal growth in early pregnancy. This latter assumption is known to be incorrect, since differences in growth according to fetal sex and maternal characteristics have been shown as early as 16-18 weeks.[60]

Work to correct errors in estimation of gestational age through statistical smoothing techniques has been done to limit its impact on birthweight-for-gestational-age percentiles, [12, 61] but inaccuracy in estimates of gestational age continue to be a concern in the creation of birthweight charts.

2.3.2 Bias at preterm gestational ages

Even with the introduction of the first birthweight-for-gestational-age chart by Lula Lubchenco in 1963, it was recognized that a potential for bias existed at preterm ages.[8] It was speculated that the pathological process responsible for preterm births may also affect fetal growth, leading to a distribution of weights in preterm newborns that was systematically lower than that of the cohort of healthy (in utero) fetuses of similar gestational age. As a result, the weight percentiles of conventional birthweight charts, which were based on only the weights of preterm newborns, would be biased at preterm gestational ages.

With the development of obstetrical ultrasound and formulae to estimate fetal weight (see section 2.4.1), it became possible to confirm this hypothesis. Weiner and colleagues[19] compared the 5th, 10th, and 50th percentiles of previously published birthweight-for-gestational-age charts and charts of estimated fetal weights between 23 and 33 weeks. The observed birthweight percentiles were significantly lower than the weight percentiles predicted using ultrasound, leading the authors to conclude that preterm infants may have had suboptimal intrauterine growth. Secher and colleagues[18] estimated the fetal weights at 223 days (31 weeks, 6 days) and 258 days (36 weeks, 6 days) of over 200 pregnancies randomly sampled from a larger study population of 3888 pregnancies. They compared the distribution of intrauterine estimated fetal weights to that of the distribution of weight derived from the birthweight-for-gestational-age chart created from the larger study sample, and found that the birthweights were systematically shifted to lower values, with the discrepancy being most pronounced at the earlier gestational age, again suggesting that preterm births are smaller than their *in utero* peers. Using a study population of 5757 deliveries, Ott[17] examined the percent of newborns classified as IUGR (defined as a weight below the 10th percentile) by intrauterine and conventional birthweight charts. They found that the use of intrauterine charts resulted in a prevalence of "IUGR" that increased with decreasing gestational age, while the use of conventional charts classified a relatively constant proportion of birth as "IUGR" across different gestational age. This further

suggested that conventional birthweight charts obscure an association between preterm birth and fetal growth restriction. Marsal and colleagues[16] arrived at the same conclusion when their ultrasound estimated fetal weight standard was applied to a population of 8663 newborns in a Swedish population. Finally, similar results were also observed in a Swiss population when an intrauterine reference was used to classify the weights of 2406 preterm births.[62] As shown in Figure 2-3, a disproportionate amount of preterm newborns had birthweights below the lower percentiles (here, the 3rd percentile) of the intrauterine weight chart.

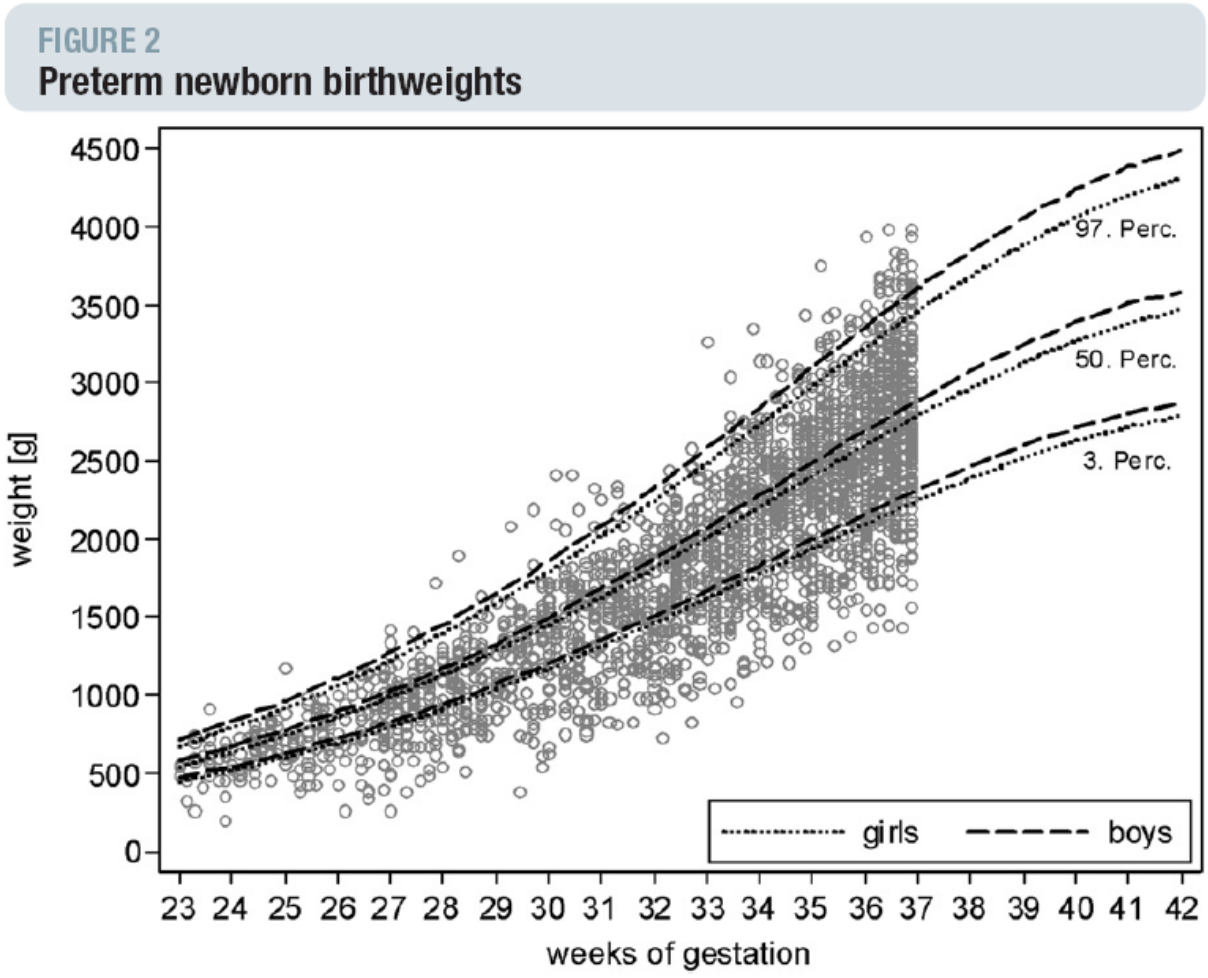


Figure 2-3. Preterm newborn birthweights compared with ultrasound estimated fetal weight reference.

Reprinted from the American Journal of Obstetrics and Gynecology "Newborn weight charts underestimate the incidence of low birthweight in preterm infants" Vol 199, page139.e3 Copyright (2008), with permission from Elsevier.

While the previous studies examined the association between preterm birth and birthweights of all newborns, Zeitlin and colleagues[63] separated preterm births according to mechanism of birth, distinguishing infants born early as a result of medical intervention from spontaneous preterm births (premature rupture of membranes were classified with the spontaneous preterm births). Since the gestational age that fetuses induced at preterm ages *would have been* born at had they not been induced is unknown, only spontaneous preterm births can provide information on any causal biological relationship (whether direct or because of a common cause) between poor fetal growth and timing of birth. Using data from a large, 17 country case-control study of preterm births, they observed that spontaneous preterm births were 1.6-fold [95% CI 1.43, 1.82] more likely to be "small-for-gestational-age" (below the 10th percentile of an intrauterine-based standard) compared with term births. Similar results were found by Morken and colleagues,[64] who examined the relationship between spontaneous preterm birth and birthweight. Using data from the Swedish Medical Birth Register, they examined the risk of spontaneous preterm birth among over 1 million births. Using a weight-for-gestational-reference based on the weights of the intrauterine population (estimated through ultrasound), a birthweight Z-score was calculated for each infant. Infants whose birthweights were more than 3 standard deviations below the expected mean were found to be 3.1-fold [95% CI 2.6, 3.6] more likely to be spontaneous preterm births, while infants 2.1-3.0 standard deviations below the expected mean were 1.2-fold more likely to be spontaneous preterm births [95% CI 1.1, 1.2]. Interestingly, infants with a birthweight above the expected mean were also significantly more likely to be spontaneous preterm births (OR spontaneous preterm birth = 1.6 for infants with a birthweight 2-2.9 SD above average, [95% CI 1.5, 1.7]), suggesting that both intrauterine growth restriction and fetal overgrowth may be associated with early parturition.

The majority of evidence supporting a link between fetal growth restriction and preterm birth is based on a comparison of weights of preterm newborns with ultrasound estimated fetal weight standards. A study by Hediger [65] provided evidence for this association using a prospective study design. Estimated fetal weights (EFW) were obtained at 32-weeks in a cohort of 290 pregnancies, sampled randomly from participants of a larger research study. The 32-week EFW of infants subsequently born preterm were found to be over 120 grams lighter than the 32-week EFWs of infants subsequently born at term. When stratified according to types of preterm births, this difference was found to be largest for infants born preterm due as a result of

medical/obstetrical indications (an average of 260 g lighter, which is unsurprising given that intrauterine growth restriction is a common indication for preterm delivery), but differences were also seen with infants born preterm due to premature rupture of membranes (an average of 169 grams) and preterm labour following failed tocolysis (an average of 137 grams).

Hediger's study[65] found that the 32-week EFW of infants subsequently preterm were already significantly smaller than those subsequently born at term. However, because of the small sample sizes of the total cohort (n=290), the number of preterm births that these conclusions were based on was extremely small: there were only 8 infants included in the "preterm birth due to medical/obstetrical indications" group, 16 born preterm due to premature rupture of membranes, and 22 born due to preterm labour. A recent study by Zhang [66] has re-examined these associations using a considerably larger population of 3,360 low-risk pregnancies who had participated in a randomized controlled trial designed to evaluate the impact of routine second and third trimester ultrasounds on perintal morbidity and mortality(the Routine Antenatal Diagnostic Imaging with Ultrasound (RADIUS trial).[67] They used a competing-risks model to examine whether fetal growth restriction (as assessed by an adjusted Z-score of the difference in weight between a second trimester ultrasound and birth) was associated with spontaneous preterm birth, which allowed the competing-risks of birth due to premature rupture of membranes and birth due to medical intervention to be taken into account. The probabilities of spontaneous birth were compared between women with normal, slow, and fast fetal growth (growth Z-scores of 0, -2, and +2, respectively). No significant differences were observed between these three groups, leading the authors to conclude that in low-risk pregnancies, intrauterine growth restriction is not associated with spontaneous onset of labour.

Further work to clarify the relationship between fetal growth restriction and spontaneous preterm birth is needed to clarify these apparently contradictory results. Nevertheless, since infants born preterm due to medical intervention make up a considerable percent of preterm births[68] and fetal growth restriction is one of the most common indications for preterm delivery,[69] there is strong evidence for the bias at preterm ages in conventional birthweight charts (since they are based on the weights of all births, not just spontaneous preterm birth).

2.3.3 Assessment of size instead of growth

Conventional birthweight-for-gestational-age charts may be biased at preterm ages because of the non-representative weights of preterm newborns, but birthweight charts additionally have a second, perhaps more fundamental flaw. Although commonly referred to as "fetal growth charts", the name is somewhat of a misnomer. The term "growth" implies a change in size between two time points. Since birthweight charts are constructed from weights obtained at a single time point for each fetus (birth), the curve obtained by joining the percentiles of attained weights at each gestational age cannot be assumed to reflect the trajectory of intrauterine growth.[15, 70] In fact, the non-representative weights of preterm newborns means that the "fetal growth curve" obtained from conventional charts is likely too steep at preterm ages. The slope of the curve between 32 and 37 weeks, for example, reflects the contributions of two distinct phenomena: 1) the growth of the fetus during this period, and 2) an increased representativeness of the weights of newborns compared to those of ongoing pregnancies. As the distribution of newborn weights transitions from being systematically lower than the weight distribution of ongoing pregnancies (at preterm ages) to being representative of the total cohort (at term), the curve obtained by joining the cross-sectional percentiles will incorporate this upward shift. Removing the effect of this bias from birthweight-for-gestational-age charts could be speculated to result in a true fetal growth curve that is less sigmoid-shaped, and more linear in nature.

The use of birthweight-for-gestational-age charts to identify individual infants with intrauterine growth restriction is likewise flawed.[15, 71] Although the pathological process of interest is a suboptimal rate of *growth*, the assessment of cross-sectional *size* is typically used as a proxy. Thus, although some infants may be in the smallest ten percent of their peers (small-for-gestational-age) as a result of restricted intrauterine development, others in the smallest ten percent will be small, but healthy and growing steadily. Although all of these infants could be correctly classified as SGA, only the former are truly growth restricted, an important (and often overlooked) distinction for the appropriate study and identification of high-risk infants. Conversely, the growth of some infants may have been restricted *in utero*, putting them at increased risk of adverse perinatal outcomes, but their absolute weight may not be in the lowest 10 percent of the population.

Danielian and colleagues explored the need to differentiate between size and growth when identifying adverse perinatal outcomes.[72] The pregnancy outcomes of a cohort of 197 unselected pregnancies were examined, including the presence of an abnormal cardiotocograph pattern (CTG) in labour, meconium staining in the amniotic fluid, fetal blood pH, need for operative delivery, apgar scores, NICU admission, and need for neonatal intubation. Infants were classified according to their size (based on a conventional birthweight-for-gestational-age percentiles) and growth (based on a percent difference between actual birthweight and expected birthweight using an extrapolation of a third trimester estimated fetal weight, with a more than 5% drop being considered as poor growth.) Discriminant analysis was used to relate the percent change in birthweight or birthweight percentile to perinatal outcomes. A >5% drop from expected birthweight was found to be significantly associated with an abnormal CTG during labour and need for operative distress, while birthweight-for-gestational-age percentile was found to be significantly associated with abnormal CTG only. Based on these results, the authors concluded that a poor pattern of fetal growth was as important as size *per se* in identifying risk for adverse perinatal outcomes.

Patterson and colleagues also compared perinatal outcomes of 355 infants classified by size as well as growth.[73] They estimated the infants' growth by measuring the ponderal index (weight in grams/length in cm^3 x 100) and midarm circumference to head circumference ratio, classifying infants with either a ponderal index or midarm circumference:head circumference below the 10^{th} percentile as "thin", and infants with values between the 10^{th} and 90^{th} percentiles for both anthropometric characteristics as "normal". Infants were also classified as being small-for-gestational-age (SGA, a birthweight below the 10^{th} percentile) or non-SGA (a birthweight above the 10^{th} percentile). The risk of adverse perinatal outcome, defined as an operative delivery for fetal distress, a 5-minute apgar score<7, meconium aspiration, polycythemia, or hypoglycaemia, was calculated according to a cross-tabluation of thin/normal and SGA/non-SGA status. As expected, the highest risk of adverse outcome was observed among infants that were both small and thin (5/11 infants, 46%). Twenty-two (22) percent of infants that were thin, but not small, were observed to have adverse outcomes (13/59 infants), while none of the infants that were small, but not thin had any adverse outcomes (0/9). The author concluded that fetal growth restriction, as established by a poor nutritional status at birth, was more important than absolute size in predicting adverse perinatal outcomes.

Although the importance of differentiating between small fetal size and poor intrauterine growth is well accepted,[15] the use of birthweight-for-gestational-age charts in the study of fetal "growth" remains widespread. Lack of routine data on third trimester estimated fetal weights and lack of an accepted classification scheme for fetal growth likely contribute to the ongoing use of size as a proxy for growth, and future work in this area is needed to change current practice.

2.4 Alternatives to conventional birthweight-for-gestational-age charts

2.4.1 Estimated fetal weight standards

To resolve the bias in conventional birthweight charts caused by the non-representative weights preterm births, information on the weights of all ongoing pregnancies and births at a given gestational age is needed. The introduction of ultrasonographic imaging into obstetrics in the late 1950s and 1960s[74] allowed the development of formulae that attempted to estimate the weight of the fetus by combining different fetal biometric measurements(Table 2-1). Early formulae predicting weight relied only on fetal biometric measurements of the abdomen, or abdomen and biparietal diameter to estimate fetal weight, such as those of Campbell and Wilkin [75](Abdominal circumference (AC) only), Shepard [76, 77] (AC and biparietal diameter (BPD)), or Eik-Nes [78] (abdominal diameter (AD) and BPD). Later formulae[79-82] additionally incorporated information on femur length into the estimation of fetal weight equations, including the widely used formula of Hadlock based on femur length, head circumference, and abdominal circumference.[79] Additional work has included the development of sex-specific equations,[83] and volumetric-based (rather than regression-based) formulae,[84] but the formula of Hadlock[79] remains arguably the most commonly used formula in North America.

Table 2-1 Common formulae for the estimation of fetal weight

Authors	Year	Formula
Campbell and Wilkin [75]	1975	Log_e(weight)= -4.564 + (0.282 x AC)- (0.00331 x AC^2)
Shepard et al.[76]	1982	Log_{10} (weight)= -1.7492+0.166/(BPD) + 0.046(AC) – 2.646(AC x BPD)/1000
Eik-Nes et al.[78]	1982	Log_{10} (weight)=1.85628 x log(BPD) + (1.34008 x log(AD)) - 2.84421
Persson et al.[80]	1986	Log_{10} (weight)= 0.972 x log(BPD) + 1.743 x log(AD)+0.367 x log(FL)-2.646
Hadlock et al.[79]	1985	Log_{10} (weight)= 1.326 – (0.00326 x AC x FL) + (0.0107 x HC) + (0.0438 x AC) + (0.158 x FL)
Hadlock et al.[84]	1985	Log_{10} (weight)= 1.3596 - (0.00326 x AC x FL) + (0.00064 x HC) + (0.00061 x BPD X AC) + (0.0424 x AC) + (0.174 x FL)
Rose et al.[81]	1987	Log_e(weight) = 0.143(BPD + AD + FL) + 4.198
Combs et al.[84]	1993	weight=(0.23718 x AC^2 x FL)+ (0.03312 x HC^3)

AC=abdominal circumference
ATD= abdominal diameter
BPD= biparietal diameter
FL= femur length
HC= head circumference

The measurement error associated with estimation of fetal weight has long been recognized. A systematic literature review has recently summarized the error associated with different formulae, as reported by either the formulae developers or other groups, in estimated fetal weight validation studies .[85] Validation studies were divided according to the study populations in which they were performed: either “normal” clinical populations, low birthweight populations (ranging from weights of 478 to 3216 grams), or high birthweight populations (4000 grams or above). In normal clinical populations, the formulae of Hadlock [79] were found to give the most consistent results, with minimal systematic error, but random error of roughly 10% for 1 standard deviation. In general, though, no single formula emerged as better than others, and all were deemed to have an unacceptably high degree of random error. The volumetric formula of Combs, [84] with a mean systematic error of 0.1%, and random error of 9.1% was deemed to be a promising alternative, but required further study. In low birthweight populations, results were highly variable, both in term of systematic error as well as random error, though the small sample sizes of these studies could also explain the variability of results. Formulae with 3 biometric measurements (e.g. AC, FL, and HC instead of only AC and HC) appeared to have a somewhat lower systematic error, but random error was still upwards of 10%. In high birthweight populations, a trend of systematic underestimation was apparent. The author of this

systematic review concluded that “No consistently superior method has emerged... Efforts must be made to minimize this variability if EFW is to be clinically useful” (p.80). He reviewed the literature on possible sources of error, including inter- and intra-rater variability, ultrasonographic imaging, and validity of the formulae themselves, and concluded that all areas should be improved to decrease the measurement error associated with EFW.

Of note, the quality of many EFW validation studies themselves is likely a barrier to our understanding of estimation of fetal weight. Most studies had small sample sizes (generally 100-200 across all birthweights in general clinical populations and fewer than 120 for low birthweight populations (as few as 13 fetuses in one recent study[86])).[85] An exception to this is the study of Kurmanavicius ,[87] which examined estimated fetal weights and birthweights in 5612 women. However, a difference of up to 7 days between time of EFW and delivery was allowed in this study. Since the fetus is gaining 20-25 grams per day during this time,[88] this difference could very likely introduce a meaningful degree of error to the validation process. A time difference of up to one week between ultrasound and delivery is not unique to this study,[86] and likely further confuses our understanding of EFW formulae validity.

Despite the random error in estimation of fetal weight, reference charts of estimated fetal weights-for-gestational-age have been produced.[16, 89-94] Hadlock and colleagues used their previously published formula[79] to develop a fetal weight-for-age curve based on 392 low-risk pregnancy from a predominantly middle-class, white population in the American Midwest.[91] The gestational ages of the fetuses were established from the first day of the last normal menstrual period confirmed with early ultrasound, and a single subsequent ultrasound examination was performed to estimate fetal weight. Having only a single weight measurement from each fetus ensured that the assumption of independence required for their log-linear regression model of fetal weight according to gestational age was met, but also meant that the reference chart reflected only cross-sectional weight measurements, rather than longitudinal changes in weight (growth) for a given fetus. Nevertheless, despite its small sample size and use of suboptimal methodology, Hadlock’s standard remains one of the most commonly used fetal weight standards in North America.

Marsal and colleagues created a fetal weight-for-gestational age reference for a Swedish population.[16] Gestational ages for the pregnancies, all low-risk research study volunteers, were

confirmed through early ultrasound, and estimates of fetal weight were calculated using the formula of Persson. [80] Though the study was based only on observations from 86 pregnancies, estimates of fetal weight were obtained every 3 weeks during pregnancy, providing up to 11 measurements per fetus. Unlike cross-sectional standards such as Hadlock's,[91] this standard is therefore able to describe longitudinal changes throughout pregnancy for individual fetuses. The correlation between multiple observations from the same fetus (that is, intra-individual correlation), however, was not taken into account when establishing the growth curve, likely meaning that the variance in their reference values was underestimated.[95] Though based on a small sample size, this reference also remains in common use.[42, 64, 96-98]

Gallivan and colleagues[90] obtained biweekly estimates of fetal weight in 67 normal pregnancies to produce a fetal "growth" chart. Unlike the modelling approach of Marsal and colleagues,[16] which essentially ignored the serial nature of the ultrasound measurements, Gallivan fit individual regression curves to each fetus, producing 67 regression curves. These regression formulae were used to calculate predicted weights at each gestational age, and the means and standard deviation of these 67 curves used to produce reference percentiles. As outlined by Royston and Altman,[95] this approach also has several limitations . First, intra-individual variability (variability of each weight measurement around a fetus' individual best-fit growth trajectory) was not incorporated into the model. As a result, the range of percentiles in their reference chart will be inappropriately narrow (too close together). Second, modelling individual growth trajectories is statistically inefficient, with a large number of parameters estimated using a relatively small amount of data (parameters for gestational age, a quadratic term for gestational age, and intercept estimated for each fetus). The parameters estimated by this type of an approach would therefore be expected to be less stable.

Perhaps the most methodologically appropriate estimated fetal weight reference chart is that of Johnsen and colleagues.[92] In a population of 634 low-risk pregnancies in Norway, 1799 estimates of fetal weight were obtained. Fetal weight was estimated using the volumetric formula of Combs,[84] which has been reported to have a lower systematic and absolute error than other formulae such as Hadlock's.[79] In addition to having a much larger sample size than previous references, the study also used a multi-level model (mixed model) to describe the serial fetal weight measurements. This approach, previously recommended for the construction of fetal weight charts,[95] correctly accounts for both inter- and intra-fetus variability in growth (variability

in rates of fetal growth, and variability of individual weight measurements around a fetus' own individual trajectory, respectively), producing unbiased estimates of population distributions of fetal weight.

Authors comparing conventional birthweight charts and intrauterine charts have concluded that intrauterine charts are better able to predict adverse perinatal outcomes related to growth restriction.[96, 99, 100] Zaw and colleagues[100] classified a cohort of 1267 infants born <34 weeks according to a conventional birthweight chart and Hadlock's[91] intrauterine standard. The classification of SGA produced by Hadlock's intrauterine standard was better able to identify risk of adverse outcomes such as intraventricular haemorrhage and respiratory distress than the classification of SGA produced by the conventional standard, though not the risk of other outcomes such as necrotising enterocolitis or retinopathy of prematurity. The relatively small number of adverse events, however, resulted in wide confidence intervals that prevented clear distinctions between the two standards. Cooke [96] examined a database containing 25 years of preterm admissions to a UK hospital. Nearly 8000 infants born at 34 weeks or younger were assigned a birthweight Z-score from both an intrauterine and birthweight standard. The odds ratios of adverse outcomes such as necrotising enterocolitis, septicaemia, periventricular haemorrhage, and neonatal mortality among infants 3, 2-3, and 1-2 standard deviations below average were calculated. Although again, the overlap in confidence intervals prevented clear distinctions between the 2 different standards, the authors concluded that the intrauterine growth standards gave a better indication of the incidence of fetal growth restriction among preterm infants and its role in adverse neonatal outcomes.

While reference charts based on intrauterine weights help resolve the bias in conventional charts at preterm ages, at term ages, they will have limitations of their own. The error in estimated fetal weight, combined with the potential for non-representativeness due to small sample sizes of research study participants, means that intrauterine charts will likely be less accurate than birthweight charts. This led to the proposal that "hybrid" charts should be used,[101, 102] with ultrasound estimates of fetal weight used to establish percentiles prior to 37 weeks, and birthweights of newborns used after and including 37 weeks. Alternatively, a hybrid chart constructed from multiple modalities has also been proposed. Mongelli [103] attempted to combine weight-for-age estimates obtained from three different sources: ultrasound estimated fetal weights, birthweights, and weight estimated through MRI (magnetic resonance imaging).

Median values from a previously published intrauterine weight-for-age reference were expressed as a fraction of weight at 280 days (40 weeks) and modelled to obtain coefficients describing the effects of gestational age on percent of 280-day weight (described by the author as a "fractional growth curve"). Median birthweights from a UK hospital obstetrical database were also expressed as a fraction of weight at 280 days. A "correction factor" was first applied to the weights of preterm births to account for their lower weight-for-ages, based on the discrepancy between the 32-week estimated fetal weights of infants subsequently born at term versus the 32-week EFW of infants subsequently born preterm. Finally, median values from a previously published MRI standard which estimated fetal volume in 18 pregnancies was also expressed as a percent of weight at 280 day and converted to a "fractional growth curve". The coefficients from all three "fractional growth curves" were then averaged to obtain a combined estimate of the effects of gestational age on fractional growth. A new reference was created by applying the combined fractional growth curve coefficients to the median weight at 280 days in the UK hospital database, and applying a coefficient of variation (CV) of 12% (that observed in the UK database at 40 weeks) to all gestational ages. Although this approach could potentially reduce the bias in conventional birthweight charts at preterm ages, the small sample sizes, crude statistical methods and assigning of equal weight to highly reliable (birthweights) and unreliable (estimates from MRI in a population of 18 women) weight observations make the representativeness of this specific reference questionable.

2.4.2 Customized birth weight percentiles

Conventional birthweight percentiles are unable to differentiate between infants who are small due to growth restriction and infants who are small, but healthy and growing well. Intuitively, many of these "small-but-healthy" infants would be expected to be born to mothers who were also small: observations from everyday life tell us that small women tend to have small offspring. In the early 1990s, Jason Gardosi and colleagues attempted to formally incorporate this intuition into the assessment of fetal growth.[104] Their goal was to incorporate information on maternal characteristics believed to have a physiological influence on fetal growth (such as maternal height, ethnicity, pre-pregnancy weight, and parity) into the calculation of birthweight percentiles. A multiple linear regression model predicting birthweight among 4179 term singleton births in a United Kingdom (UK) population was generated using maternal height, pre-

pregnancy weight, parity, ethnicity, fetal sex, and gestational age as independent variables. The coefficients from this model were then use to calculate a physiologically "optimal" birthweight for each infant. Thus, the offspring of a nulliparous women born at 40 weeks of gestation would have an "optimal" weight that was expected to be 53 grams lighter than that of a woman delivering her second offspring, while the 40 week old offspring of a woman taller than 170 cm would have an "optimal" weight 69 grams heavier than a woman 161 to 170 cm tall. An 80% reference range (i.e. 10^{th} to 90^{th} percentiles) was calculated for each infant's "optimal" weight from the model's residual variance. Infants whose observed birthweight was below the 10^{th} percentile of their own individual "optimal" weight predicted by the regression model were classified as "small-for-gestational-age" by the customized standard.

The methodology for these "customized" percentiles was further developed in a subsequent publication.[105] The values of the coefficients for maternal characteristics were re-established using a larger sample size of 38 114 singleton term births from the same UK population through stepwise regression. Further, a "proportionality" formula was developed to better estimate the "optimal" birthweights at early gestational ages. Using Hadlock's previously published intrauterine (ultrasound) standard,[91] the median weight at each day of gestation between 24 and 42 completed weeks (168 to 294 days) was expressed as a proportion of the median weight at 40 weeks (280 days). These values were then modelled using linear regression to develop a "proportionality" formula which describes the shape of the growth curve by which the weight of an infant born at 40 weeks is reached. Applying a constant coefficient of variation of 11% across all gestational ages (the observed coefficient of variation at term in this population) allowed this process to be repeated to obtain proportionality formulae for the 10^{th} and 90^{th} percentiles (coefficient of variation x 1.28 standard deviations). The optimal weight and 80% coverage limit for infants born at any gestational age could then be calculated, allowing a "customized" percentile to be generated by comparing the infants' observed birthweight to their predicted physiologically optimal birthweight.

Numerous studies have evaluated the ability of customized birthweight percentiles to identify infants at increased risk of adverse perinatal outcomes due to growth restriction.[106-113] Mongelli & Gardosi [107] recruited a group of 267 low-risk singleton pregnancies with normal clinical outcomes (defined as infants born at term, free of congenital anomalies, a 5 minute Apgar of at least 7, an umbilical cord pH>7.2, an umbilical cord base deficit greater than 8, and not requiring

neonatal intensive care unit admission). Serial ultrasounds were obtained (every 2-3 weeks after 26 weeks of gestation), and each infant was classified as SGA (crossing below the 10th percentile) according to both conventional birthweight-for-gestational-age percentiles and customized birthweight percentiles. Of the infants classified as SGA by the conventional standard (meaning that they had crossed below the 10th percentile of the conventional standard at some point during gestation), 27.5% were re-classified as 'non-SGA' by the customized standard. Conversely, of those identified as non-SGA by the conventional standard, only 2.3% were re-classified as SGA by the customized standard. Based on these results, the authors concluded that the use of customized percentiles reduces the false-positive rate for the diagnosis of intrauterine growth restriction in low-risk populations.

A second study, performed in the Netherlands, examined customized percentiles in a high-risk population.[108] The birthweights of 217 infants from pregnancies considered to be at increased risk of uteroplacental insufficiency (because of maternal hypertension, smoking, history of intrauterine growth restriction, or advanced maternal age) were classified as SGA or non-SGA according to both the Dutch birthweight-for-gestational-age chart as well as the customized birthweight chart developed by Gardosi.[104] The incidence and odds ratios for adverse perinatal outcomes such as artificial ventilation, admission to the NICU, low Apgar score, and caesarean section for fetal distress among SGA and non-SGA infants were calculated using each of the charts. The odds ratios for adverse perinatal events among infants classified as SGA by the customized standard were consistently higher than the odds ratios obtained with the conventional birthweight-for-gestational-age chart, suggesting that the customized percentiles were a better tool to identify infants at increased risk of adverse outcomes due to growth restriction. Due to the small sample size, however, the role of chance differences could not be ruled out for most of the outcomes studied.

From these small, earlier evaluations of customized percentiles, several large, population-based studies were performed. Clausson and colleagues[106] examined the risk of stillbirth, neonatal death, and low apgar among over 325,000 births recorded in the Swedish Medical Birth Register between 1992- 1995. Infants classified as SGA by the customized standard only (i.e. SGA by the customized standard, non-SGA by the conventional standard) were found to have a 6.1-fold increased risk of stillbirth compared to infants classified as non-SGA by both standards [95% CI

5.0, 7.5], while infants classified as SGA by the conventional standard only were not found to be at higher risk than infants classified as non-SGA by both standards (OR= 1.2 [0.8-1.9]). The trend in odds ratios for neonatal death was similar, with an odds ratio of 4.1 [2.5, 6.6] among infants classified as SGA by the customized standard, compared to 0.9 [0.3, 2.3] among infants classified as SGA by the conventional standard. Although the difference in odds ratio for low (≤3) 5 minute Apgar obtained from the customized and conventional standards was lower than those for mortality, the odds ratio produced by the customized standard was still statistically significantly higher than that of the conventional standard (2.2 [1.9, 2.7] and 1.2 [0.9, 1.5], respectively).

In a multi-site study performed in France, the birthweights of 56,606 infants were classified according the conventional and customized birthweight standards.[114] Infants classified as 'SGA by the customized standard only' were found to be at 2.6-fold increased risk of perinatal death [95% CI 1.62, 4.15] compared to infants classified as non-SGA by both standards, while infants classified as 'SGA by the conventional standard alone' had only a 1.08-fold increased risk [0.43, 2.51]. The odds ratios for Caesarean section and admission to the NICU were also modestly higher among infants classified as 'SGA by the customized standard only' than those classified as 'SGA by the conventional standard only', but the overlap in confidence intervals does not rule out that these differences were due to chance. The odds of an apgar score below 7 were modestly higher among infants classified as 'SGA by the conventional standard', but again, confidence intervals had a large degree of overlap.

In a large, hospital-based cohort from New Zealand, McCowan and colleagues examined the risk of morbidity and mortality in a cohort of infants suspected of being SGA (n=374) as well as a general obstetrical population (n=12,879).[111] In the cohort of suspected SGA pregnancies, the point estimate for relative risk of morbidity (defined as a composite outcome of perinatal death and/or prolonged hospital stay) was found to be highest among infants classified as SGA by the customized standard only (RR=8.7 [95% 2.8, 27]), slightly lower point among infants classified as SGA by both the customized and conventional standards (RR= 6.6 [2.2, 20]), and lowest among infants classified as SGA by the conventional standard only (RR=1.4 [0.3, 6.4]). In the general obstetrical population, the relative risks of composite morbidity were significantly higher among infants classified as SGA by the customized standard only (RR= 3.8 [3.0, 4.8]) or both the

conventional and customized standards (4.5 [3.8, 5.3]) than the conventional standard (RR=0.4 [0.1, 0.9]). These results, combined with the results from additional outcomes such as hypoglycaemia and NICU admission, lead the authors to conclude that use of customized percentiles would likely lead to the detection of more infants at risk of perinatal morbidity and mortality than the use of conventional percentiles.

Not all studies demonstrated a benefit of customized percentiles over conventional methods. Lyon and colleagues[109] examined the perinatal autopsy records of 51 stillbirths with an unexplained cause of death. They classified the stillbirths' weights using both a customized birthweight chart and an intrauterine weight-for-gestational-age chart, and examined the ability of each to identify fetal growth restriction, as defined by a brain-to-liver ratio of > 5 (severe IUGR) and >3 (moderate IUGR). The ability of the two standards to identify fetal growth restriction was found to be similar, with high sensitivities (95% for both) but low specificities (63% for the intrauterine weight-for-age chart, and 66% for the customized chart) for severe growth restriction. Most recently, research from our group at McGill[1] re-examined the Swedish Medical Birth Register database used in the earlier publication of Clausson.[106] As in the earlier publication of Clausson, they found that infants classified as SGA by the customized standard (ie. classified as non-SGA by the conventional birthweight standard) only were at increased risk of adverse outcomes (stillbirth, neonatal mortality, and low 5 minute apgar score) compared to infants classified as non-SGA by both standards (OR=7.8, [6.9, 8.9]). However, it was also noted that infants classified as SGA by the customized standard only were also more likely to be born preterm. Over 16% of infants identified as SGA by the customized standard only were born before 37 weeks, as opposed to only 7% of infants identified as SGA by both standards, 3.4% of infants identified as SGA by the conventional standard only, and 4.2% of infants identified as non-SGA by both standards. Once this difference in gestational age was controlled for, the high odds ratio of adverse outcomes among infants classified as SGA by the customized standard was greatly reduced, to 2.4 [2.1, 2.8], suggesting that differences in gestational age between the groups were responsible for the previously observed high odds ratios.

Nevertheless, the use of customized percentiles has been steadily gaining popularity in the clinical and research communities. Practice guidelines of the British Royal College of Obstetricians and Gynaecologists have recommended the use of customized percentiles for the investigation of the small-for-gestational-age fetus,[115] and a recent editorial in the American

Journal of Obstetrics and Gynecology concluded that "…it would seem to be an appropriate time for American obstetricians to adopt the use of customized fetal growth standards."(p.21).[116] A program to calculate customized birthweight percentiles has been made available on the internet (http://www.gestation.net/birthweight_centiles/birthweight_centiles.htm) (Accessed Sept 12, 2008), and there is ongoing research to establish customization coefficients for different populations, including Britain,[105] Australia,[117] New Zealand,[118] France,[114] Sweden,[106] and the United States .[119]

2.4.3 Assessments of growth

Ultrasound estimated fetal weight charts and customized charts may help resolve the bias in preterm ages in conventional birthweight-for-gestational-age charts, but they remain classifications of fetal *size* rather than fetal *growth*. While any past fetal weight-for-gestational-age percentiles are likely used informally by obstetricians to "eyeball" a fetal growth trajectory, there have only been a limited number of attempts to develop formal standards that assess fetal *growth*. In the clinical literature, a variety of simple measures have been proposed.[88, 120-126] Owen and colleagues[88] published a reference for fetal growth based on serial ultrasounds in 274 low-risk pregnancies. Change in fetal biometric measurements (including estimated fetal weight) between 28-day periods were averaged to produce a mean and standard deviation of grams gained per day for each week of gestation between 26 and 40 weeks. Reference values for the mean and standard deviation of percent change in weight per day throughout this period were also reported. Chang and colleagues[120] used reference values from previously published intrauterine weight-for-gestational-age charts to calculate an estimated fetal weight Z-score for each ultrasound estimate of fetal weight, then expressed fetal growth as the change in Z-scores between two time points. Smith-Bindman and colleagues[123] established a measure of growth by first calculating the average grams gained per day in their population of 321 women who had received 2 or more ultrasounds between 13 and 38 weeks. Each infant's observed growth was then expressed in relation to the average growth from the midpoint of the growth interval, and the distribution of values in the population were then converted to "growth" percentiles. Deter and colleagues[121, 122] proposed an assessment of fetal growth based on a mathematical equation expressing the natural log of weight as a function of gestational age and a pre-established constant. The regression equation was used, in essence, to extrapolate from a weight measurement earlier in pregnancy to obtain an expected weight for the current age. Actual

current weight was then expressed as a percent difference of predicted weight to quantify "growth". Finally, Salomon and colleagues[126] developed a regression model of "fetal growth potential". The difference in fetal biometric measurements between two time points was first expressed as an average change per day. This average daily growth was then included in a stepwise regression model along with current biometric measurements and gestational age to develop a formula expressing the "fetal growth potential" of an infant.

If the goal is to describe average rates of growth at a population-level, simple measures such as a change in fetal weight Z-score or average grams gained per day may provide reasonable estimates. However, if the goal is to assess the growth of a *specific* fetus and identify if *its* growth is sub-optimal, these approaches have several shortcomings.[95] They are unable to incorporate the concept of regression to the mean, whereby a fetus at an extreme percentile value will be more likely to be closer to the population average on subsequent measurements, and they generally impose an unrealistic assumption of linearity in fetal growth trajectories. Most importantly, they fail to account for the sources of variability in serial ultrasound data, which is present both between fetuses (as fetuses grow at different rates) as well as within fetuses (caused by measurement error in a given estimate of EFW and variability of a fetus' growth around its own smoothed growth trajectory). To determine if a given fetus' growth may be deviating from healthy patterns, it is important to have an understanding of the amount of *non-pathological* variability that each estimated fetal weight can be expected to exhibit from its overall growth trajectory. Such shortcomings are perhaps why the predictive ability of simple clinical measures of growth velocity have not been demonstrated to be dramatically better than conventional birthweight references in identifying adverse perinatal outcomes associated with fetal growth restriction.[123, 127-131] The small sample sizes of most studies evaluating measures of growth (generally several hundred infants) may be a factor in the lack of strong evidence supporting the use of these measures.

In the statistical literature, the need for models which better reflect the longitudinal nature of fetal growth has been recognized. More sophisticated models, which make use of multi-level modelling approaches, have been proposed.[132-134] These models allow the construction of conditional growth percentiles, whereby the percentile assigned to a fetus at a current ultrasound assessment is calculated taking into account (conditional on) its weights from past measurements.[134] A fetus whose current weight is precisely as would be expected given past

measurements would be assigned the 50th conditional growth percentile, while a fetus whose current weight was below an 80% coverage limit of the EFW expected given past measurements would be assigned a conditional growth percentile below the 10th percentile, and considered growth restricted. By explicitly incorporating information on past weight into the assessment of current weight, conditional percentiles are able to assess fetal *growth* instead of simply *size*.

Reference values of conditional fetal growth percentiles have been published for a Norwegian population,[92] based on a sample of 634 low-risk pregnancies , and a UK population, based on a sample of 274 low-risk pregnancies.[135] The authors of the UK study examined the conditional percentiles in relation to anthropometric characteristics at birth,[110] but the extent to which conditional growth percentiles are able to improve identification of fetuses at increased risk of important adverse perinatal outcomes due to growth restriction remains to be established.

3. Research objectives

The objectives of this thesis were:

1) To evaluate the potential for bias arising from the use of conventional birthweight-for-gestational-age percentiles in epidemiologic studies of fetal growth restriction and,

2) To evaluate two alternatives to conventional birthweight-for-gestational-age percentiles: customized birthweight percentiles and conditional fetal growth percentiles.

4. 'Small-for-gestational-age' as an outcome in epidemiological research

4.1 Preamble to manuscript 1

Birth weight percentile charts are most commonly used to monitor fetal growth and screen for fetal growth restriction in the clinical setting, but equally important is their use as an outcome in etiologic studies of fetal growth restriction. "Small-for-gestational-age", typically defined as a birth weight below the population 10th percentile of weight for gestational age, is arguably one of the most commonly used outcomes in perinatal epidemiology. However, the validity of using "small-for-gestational-age" as an exposure or outcome definition given the known shortcomings of conventional birthweight charts has received little attention. This chapter provides a critical appraisal of the use of "small-for-gestational-age" in epidemiologic studies of fetal growth restriction. The manuscript, entitled "The missing data problem in birth weight percentiles and thresholds for 'small-for-gestational-age'", was published in the American Journal of Epidemiology (2008; 167:786-792). A reprint of this article, as well as an Invited Commentary written by Dr. Nigel Paneth and our response, is included in Appendix G.

4.2 Title Page & Footnotes

Title: The missing data problem in birth weight percentiles and thresholds for "small-for-gestational-age"

Authors: Jennifer A Hutcheon[1] and Robert W Platt[1,2]

Affiliations: McGill University Faculty of Medicine, Departments of Epidemiology & Biostatistics[1] and Pediatrics[2]

Corresponding author:
Jennifer Hutcheon
The Montreal Children's Hospital Research Institute
4060 Rue Sainte Catherine Ouest #205
Westmount, QC, Canada
H3Z 2Z3
Tel: 514.934.1934 x-23283
Fax: 514.412.4331
jennifer.hutcheon@mail.mcgill.ca

Address for reprints:
Jennifer Hutcheon
The Montreal Children's Hospital Research Institute
4060 Rue Sainte Catherine Ouest #205
Westmount, QC, Canada
H3Z 2Z3
jennifer.hutcheon@mail.mcgill.ca

Running title: The missing data in birth weight percentiles

MeSH keywords: Bias, Epidemiologic; Infant, Small for Gestational Age; Intrauterine Growth Retardation; Reference Values
Abbreviations: AGA, appropriate-for-gestational-age; SGA, small-for-gestational-age

Acknowledgments: We thank Dr. Maala Bhatt for comments on earlier versions of the manuscript and Dr. Jean-François Boivin for helpful suggestions following a seminar presentation of this work. JAH is supported by a Canadian Institutes of Health Research (CIHR) Doctoral Research Award, RWP holds a Chercheur-Boursier from the Fonds de la Recherche en Santé Quebec (FRSQ).

4.3 Abstract

Weight-for-gestational-age charts and definitions of ‘‘small-for-gestational-age’’ based on the distribution of livebirths at a given gestational age have conventionally been used to identify infants whose fetal growth is poor. However, references based on the weights of only livebirths have serious shortcomings at preterm ages due to missing data on the weights of fetuses still in utero, and these missing data introduce considerable bias to etiologic studies of fetal growth restriction. Application of standard epidemiologic approaches for missing data is needed to help produce perinatal weight percentiles that provide unbiased assessment of fetal growth and risks of small-for-gestational-age.

4.4 Introduction

Since their first publication by Lubchenco et al.[8] over 40 years ago, weight-for-gestational-age charts[9, 11-13] have been a cornerstone of screening for infants whose intrauterine development is poor. In these charts, the weight distributions of livebirths at each week of gestation are converted to percentiles, and any infant whose weight falls below a certain statistical threshold of the population, typically the 10th percentile, is labeled as being ''small-for-gestational-age'' (SGA) and is considered to be at increased risk of perinatal morbidity and mortality.[4, 7, 136] SGA, an anthropometric characteristic that does not necessarily have any adverse health implications, is therefore commonly (but not necessarily appropriately) used as a proxy for the pathologic outcomes believed to be associated with an inadequate rate of fetal growth.[15] Weight-for-age charts are nevertheless considered an improvement over low and very low birth weight cutoffs because they differentiate between infants who are small because they are born early in gestation and infants born later but small relative to their peers.[37]

In addition to clinical use for the identification of high risk infants, weight-for-gestational-age charts and their resulting thresholds to define SGA are frequently used in epidemiologic studies of fetal (intrauterine) growth restriction.[46-48, 137-139] Because fetal growth restriction is typically not measurable in population-based data, the majority of research to identify risk factors for fetal growth restriction consists of comparisons of the risks of SGA among exposed and unexposed groups of infants. Although case definitions for SGA or ''appropriate-for-gestational-age'' (AGA) established by using conventional weight-for-gestational-age charts are well accepted in perinatal epidemiology, their validity according to general epidemiologic principles has rarely been considered.

The purpose of this article is to argue that size-for-gestational-age charts created from the weight distributions of livebirths have serious shortcomings due to missing data on the weights of fetuses that remain in utero at each gestational age. We will demonstrate how use of a case definition for SGA produced by these charts can introduce considerable bias to estimates of relative risk in etiologic studies of fetal growth restriction, and we will propose that standard statistical and epidemiologic approaches to missing data could be applied to address the bias that currently exists in this field of research.

4.5 The missing data in birth weight references

In the study of fetal growth, perinatal researchers have traditionally been faced with what is, in essence, a ‘‘missing data’’ problem.[140, 141] Although the biological process of interest is the changing fetal size throughout pregnancy, this process is generally unobservable at a population level. In a cohort of conceptions followed forward in time, information on fetal size at any given week of pregnancy is readily available for only the portion of the cohort born during that week. The weights of the remainder of the conception cohort at that gestational age, the fetuses still in utero, are unavailable.[142] Although prenatal ultrasonography has enabled estimation of fetal weight prior to the time of birth,[76, 79] this information has not been incorporated into reference charts because of concerns over measurement error.[12] National weight-for-gestational-age charts[9, 11-13] therefore continue to be created from the weight distributions of only livebirths at each age and are missing the weights of fetuses in the conception cohort not yet born by the end of a given gestational week. In these charts, calculation of weight percentiles at early gestational ages is based on the weights of an extremely small fraction of the total cohort at risk (since the vast majority remain in utero at preterm ages), and, even for gestational ages as old as 36 weeks, weight data for more than 97 percent of the original cohort are still missing.[12]

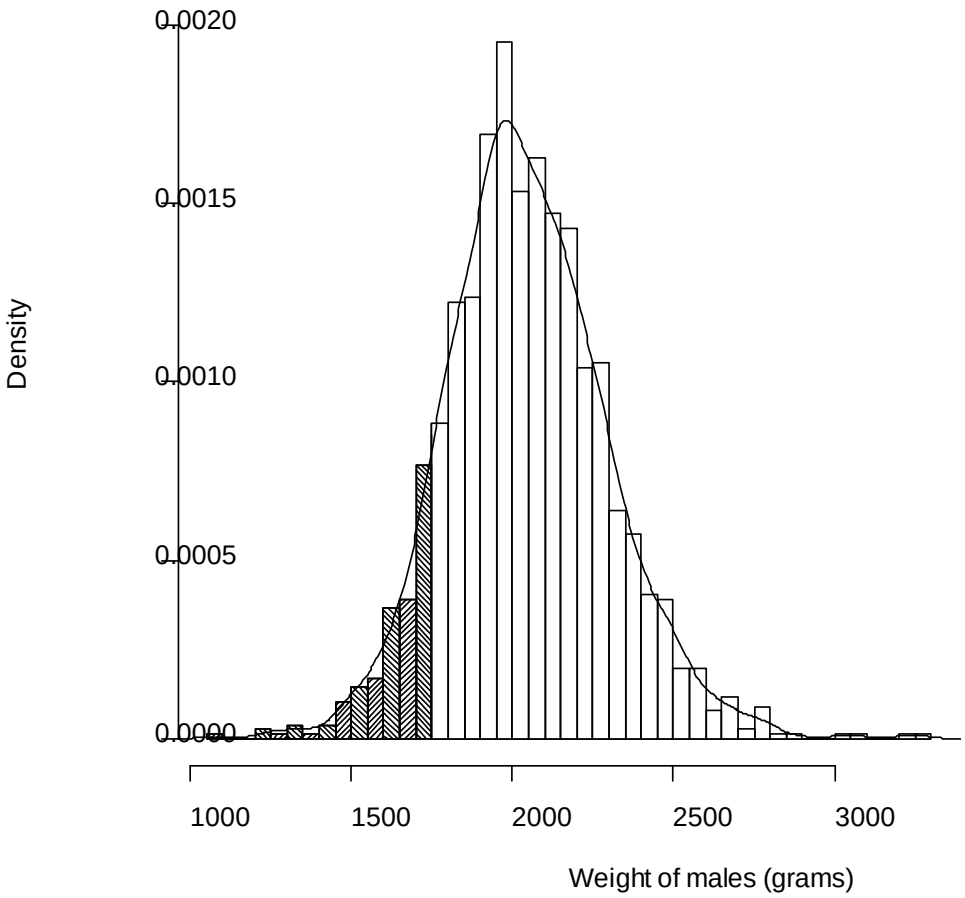
As with any missing-data situation in epidemiology, the extent to which bias will be introduced by the missing intrauterine weights will depend on the mechanism that caused the missingness.[143-145] In order for the ‘‘complete case’’ approach[146] used in conventional weight-for-gestational-age charts to be valid (the use of only those cases for whom complete data are available—i.e., data on live births), the unobserved data must be missing completely at random (MCAR). For data to be missing completely at random, they must represent a randomly selected subset of the total cohort at risk. At term, when the weight distribution of those born at a given gestational age (such as 39 weeks) is likely fairly representative of those still in utero (those who will be born at 40 weeks or later), the assumption of being ‘‘missing-completely-at-random’’ may be reasonably valid, and there will be minimal bias from the missing data on intrauterine weights. At preterm ages, however, the available weights are most likely not a random sample of the weight distribution of the total at-risk population. Intrauterine growth restriction is a common indication for medically necessary preterm birth,[68, 69] and the observation that preterm

livebirths are smaller than their in utero peers[16-19, 65, 102] has led to speculation that there may be a common cause of spontaneous preterm birth and poor growth. As a result, the observed distribution of weights at earlier gestational ages is systematically shifted to lower values than the weight distribution of the remainder of the cohort at risk at the start of that gestational week. The ''complete-case'' approach used in existing reference charts, when the missing data are clearly not missing completely at random, is therefore inappropriate.

Recognition of the ''missing data'' problem in weight-for-gestational-age charts is certainly not new. Even with the introduction of the first neonatal weight percentile charts in 1963, Lula Lubchenco warned that ''the sample has an undeterminable bias because premature birth itself is probably related to unphysiological states of variable duration in either mother or fetus. Since the weight of fetuses that remain in utero cannot be measured, the curves presented herein are submitted with these reservations . . . '' (p.793).[8] Differences between the weights of preterm livebirths and their in utero peers may be well acknowledged, but what does not appear to have widespread appreciation is the extent and impact of the bias that the missing data introduce.

The major discrepancy between intrauterine and livebirth weight distributions at preterm ages, as reported in previous publications,[16-19] is illustrated in figure 1. In the figure, the distributions of estimated fetal weights[79] of male singletons aged 32 weeks in an unselected obstetric population at the Royal Victoria Hospital, a McGill University teaching hospital in Montreal, Canada (unpublished data), are compared with a Canadian birth weight reference.[12] These ultrasound data are from the years 2001–2004 and were obtained through an institutional policy of universal 32-week ultrasound examinations. The median estimated weight of the fetuses still in utero is more than 120 g heavier than that of livebirths, while the 10th percentile (SGA) threshold of the intrauterine population is more than 300 g higher than the 10th percentile of the national birth weight reference. Similar results were obtained for female fetuses (data not shown). This discrepancy between the 10th percentile thresholds of the two distributions means that applying the national birth weight reference to the intrauterine population (which, at this age, constitutes >99.7 of the total conception cohort[12]) will not identify 10 percent of the population as SGA. Instead, since the 10th percentile of the national birth weight reference is much lower than the 10th percentile of the total cohort, the livebirth weight-based reference will identify less than 1 percent of the total cohort as SGA. That is, the SGA threshold produced by a

national birth weight reference at 32 weeks will capture the smallest 1 percent of the total cohort instead of the smallest 10 percent. Error arising from the use of a formula to estimate fetal weight will introduce some bias to estimates of the discrepancy between in- and ex-utero weight distributions; however, since most of this error is random, not systematic,[79] it is unlikely to explain a major portion of the discrepancy. The genuine discrepancy between the weights of the in- and ex-utero populations is supported by work such as Hediger et al.'s,[65] who demonstrated that the 32-week estimated weights of fetuses that were later born preterm were significantly lower than the 32-week estimated fetal weights of those that were subsequently born at term.


0.0020
0.0015
0.0010
0.0005
0.0000
Density
1000
1500
2000
2500
3000
Weight of males (grams)

4.6 Impact of biased weight percentiles on perinatal epidemiology

Birth weight references that ignore the missing data of intrauterine weights introduce considerable bias into epidemiologic studies of the etiology of fetal growth restriction. In many studies,[46-48, 137-139] the effect of potential risk factors on fetal growth restriction is evaluated by establishing the relative risk of being SGA between exposed and unexposed infants, calculated as:

Relative $risk_{SGA}$= (no. $SGA_{exposed}$/no. at-$risk_{exposed}$)/(no. $SGA_{unexposed}$/no. at-$risk_{unexposed}$) (equation1).

If the outcome, SGA, is established using a conventional reference based on the distribution of livebirths, an infant's chance of being classified as SGA will change according to his or her gestational age at birth. Consider the case of a male fetus weighing 1,650 g at 32^0 weeks of gestation (32 weeks, zero days). When the weight of this fetus is compared with the weights of all pregnancies that progressed to 32 weeks, this fetus is at approximately the 5th percentile of the population (based on figure 1 data). If all pregnancies in this cohort continued at similar relative growth rates (i.e., the rank order of weights remains unchanged) until birth at 40 weeks' gestation (term), this infant would be classified as SGA by virtue of being in the smallest 10 percent of the population of births at 40 weeks. However, if this fetus were instead born the following day, at 32^1 weeks of gestation, under existing national birth weight references[12] it would instead be assigned the 25th percentile and considered AGA. Although its weight relative to that of its peers of similar gestational stage is constant, classification of this infant as AGA or SGA will be different according to the timing of its birth. A younger gestational age at birth therefore becomes ''protective'' against being classified as SGA when based on a reference created from weights of livebirths.

The impact of gestational age at birth on the criteria for being defined as SGA becomes problematic in perinatal epidemiology because many risk factors for growth restriction (e.g., smoking, preeclampsia/pregnancy-induced hypertension, multiple births, and disadvantaged ethnicity[147-152]) have also been found to be associated with a younger gestational age at birth or increased rate of preterm birth. As a result, this leads to a differential case definition of SGA being applied to exposed and unexposed groups. Exposed infants are more likely to be born at a

younger gestational age, and, at younger gestational ages, the threshold to be identified as SGA is more stringent. At 32 weeks, for example, an infant must be among the smallest 1 percent of his or her remaining conception cohort to be labeled SGA, while, at 40 weeks, the infant need be among only the smallest 10 percent. This difference results in relatively fewer exposed infants being classified as SGA compared the unexposed group, for whom the threshold for SGA is less stringent because of older mean age at birth. As evident from equation 1, an underdiagnosis of SGAexposed infants will result in underestimation of the risk of SGA among the exposed and an underestimation or potentially even a reversal of the true measure of effect.

The amount of bias introduced because of differential misclassification of preterm SGA neonates as AGA can be quantified through a simple simulation (table 1). To begin, an estimate of the relative risk of SGA among newborns exposed to preeclampsia (compared with normotensive pregnancies) determined by using a livebirth reference was obtained from previously published research,[139] along with the mean gestational ages at birth in each exposure group. The reported unadjusted relative risk of SGA was 2.72, with a mean gestational age among the unexposed of 39.0 weeks (standard deviation, 2.3) and a mean gestational age among the exposed of 37.4 weeks (standard deviation, 3.4). These values were used to generate cohorts of 10,000 exposed and 10,000 unexposed newborns. For each gestational age, the percentage of infants whose weight was in the smallest 10 percent of the total cohort, but not of livebirths, was calculated (i.e., the percentage of SGA infants misclassified as AGA because of the use of a reference based on livebirths was established). The percentage of misclassification at each gestational age was determined by comparing the 10th percentile thresholds of a Norwegian birth weight reference[13] and a Norwegian longitudinal ultrasound reference[92] prior to 37 weeks, at which age the misclassification was zero.

The percentage of misclassifications was then used to ‘‘correct’’ the number of SGA cases at each gestational age for both exposed and unexposed groups. As expected, the number of SGA cases increased more in the exposed than in the unexposed group following the correction, since the younger mean gestational age at birth among the exposed would make them more subject to misclassification as AGA. The relative risk of SGA among the exposed was recalculated with the corrected number of SGA cases. The relative risk of SGA of 2.72 presented in the original study when a livebirth reference was used was recalculated to a relative risk of 3.24, a nontrivial

difference in effect size that creates a real possibility that true effects of exposures could be found nonsignificant or even potentially reversed because of the differential misclassification of SGA infants. For example, had a relative risk of 0.8 been found with the use of a livebirth reference, the true measure of effect would actually likely be a nearly null effect (relative risk = 0.95 based on Norwegian data, calculations not shown). Covariate adjustment for gestational age as a means to correct this problem is not appropriate, since stratification by gestational age is similar to calculating gestational-age-specific hazards with a denominator of livebirths, instead of fetuses at risk .[153]

The differential misclassification not only will affect observed measures of effect but could also create apparent, but likely spurious, biologic interactions. In a recent study, the relation between SGA birth in a first pregnancy and risk of stillbirth in a subsequent pregnancy was examined .[154] The authors reported that the risk of stillbirth in a second pregnancy increased with decreasing gestational age at birth of an SGA infant in the woman's first pregnancy (odds ratio of stillbirth after ''very preterm SGA birth'' > odds ratio after ''preterm SGA birth'' > odds ratio after ''term SGA birth'' when compared with AGA of all ages) and concluded that ''interestingly, the results in this study also reveal that SGA should be considered a heterogenous disease in terms of risk amplitude for subsequent stillbirth. A woman with a term SGA in an index pregnancy is at lower risk level than her counterpart who experiences a preterm SGA, and the greatest risk for stillbirth occurs in women with very preterm SGA'' (p. 855).[154] Before concluding that there may be effect modification in the effects of SGA on the risk of stillbirth by gestational age at birth, the potential impact of the bias from livebirth references in this study should be considered. Because those ''very preterm infants'' classified as SGA were in approximately the lowest 1 percent of their conception cohort (based on figure 1 data), whereas the SGA infants at term were in only the lowest 10 percent, it is perhaps not surprising that the more severe cases of growth restriction that consisted of the ''very preterm'' group were found to be a marker for a much greater risk of subsequent stillbirth.

TABLE 4-1. Bias introduced to relative risk of small-for-gestational age (SGA) due to misclassification of preterm SGA infants in a simulated population of 10,000 exposed and 10,000 unexposed infants, with percent of misclassification based on Norwegian population data.

Gestational age at birth (weeks)	*% SGA infants misclassified as AGA**	*Unexposed cohort (gestational age 39.0±2.3 weeks†)*			*Exposed cohort (gestational age 39.0±2.3 weeks†)*		
		No. births	Observed No. SGA births‡	Corrected No. SGA births	No. births	Observed No. SGA births§	Corrected No. SGA births
25	65.5				1	0.3	0.8
26	70.2				1	0.3	0.9
27	72.56				11	3.0	10.9
28	71.5				17	4.6	16.2
29	71.0	1	0.1	0.3	66	18.0	61.9
30	65.1	1	0.1	0.3	134	36.4	104.4
31	59.0	7	0.7	1.7	225	61.2	149.3
32	52.0	29	2.9	6.0	358	97.4	202.9
33	43.0	89	8.9	15.6	602	163.7	287.3
34	30.1	274	27.4	39.2	793	215.7	308.6
35	23.0	534	53.4	69.4	1,099	298.9	388.2
36	13.9	999	99.9	116.0	1,246	338.9	393.6
37	0	1,341	134.1	134.1	1,188	323.1	323.1
38	0	1,723	172.3	172.3	1,159	315.2	315.2
39	0	1,688	168.8	168.8	995	270.6	270.6
40	0	1,413	141.3	141.3	776	211.0	211.1
41	0	964	96.4	96.4	558	151.8	151.8
42	0	558	55.8	55.8	350	95.2	95.2
43	0	238	23.8	23.8	211	57.4	57.4
Total		9,859	985.9	1,041.07	9,790	2,662.9	3,349.46
Risk of SGA per 100			10	10.6		27.2	34.2
Relative risk of SGA#						2.72	3.24

*established by calculating the percent of infants at preterm ages below the 10th percentile of a Norwegian intrauterine weight reference[13] that were not identified as SGA by a Norwegian birth weight reference[92]; AGA, appropriate-for-gestational age

†cohort distribution truncated at 43 weeks of gestation therefore total number of births does not add to 10 000

‡number of SGA births observed with use of a live birth weight-based reference assuming risk of SGA is 10 percent among unexposed
§number of SGA births observed with use of a live birth weight-based reference if observed relative risk among exposed is 2.72
compared to unexposed cohort

4.7 Correcting the missing data bias

To correct the missing data problem in weight-for-gestational- age charts, epidemiologic methods for missing data that are consistent with the nature of the missingness should be applied.[143, 145] At preterm ages, the missing data in neonatal weight references are clearly not missing completely at random, making the current ‘‘complete-case’’ approach inappropriate. If the distribution of missing intrauterine weights were similar to that of the available birth weight data within strata of known covariates (i.e., if we were able to predict the missingness based on known covariate information), the data would be missing at random (MAR). With missing-at-random data, approaches such as multiple imputation[145] or inverse weighting[155] could be used to build references that accounted for the missing weights. However, since our ability to explain the missingness (amounting to predicting gestational age at birth) is generally agreed to be poor, even considering all known social and medical risk factors,[156] these data are likely not missing at random. The missing data in neonatal weight references are therefore likely missing not at random(MNAR), meaning that the missingness process depends on unobserved variables, and any weight-for-gestational-age reference must take this missing data mechanism into account.

A variety of attempts to address the bias from missing intrauterine weights have been proposed in the literature, but none have appropriately addressed the missing-not-at-random nature of the data. Population references based on the distribution of estimated fetal weights[16, 91, 92] represent an improvement at preterm ages[99, 100] but, later in gestation, will introduce missing data bias of their own because of missing weights for those in the population who have been born. ‘‘Hybrid’’ references, which either average the growth curves created from livebirth and intrauterine weights[103] or switch from intrauterine weight distributions to birth weight distributions at 37 weeks,[101] have also been proposed. While correct in spirit, neither of these approaches accurately reflects the portions of the population in- and ex-utero at each gestational age.

Although options for analyzing missing-not-at-random data are usually limited,[143, 145] the case of neonatal weight references represents a relatively rare situation in which external data can be incorporated to produce valid results. With missing-not-at-random data, the weight distributions will be different between those with and without missing data, even within strata of observed

covariates. Here, estimates of the weight distributions of those with missing data can be obtained from estimates of fetal weight produced by obstetrical ultrasound.[76, 79] Although such estimates have a considerable amount of random error,[85] this problem is mainly of concern for predicting weight at the individual level, not for the weight distributions of the population as a whole. Since the magnitude and direction of error in estimates of fetal weight have been reported in validation studies for fetal weight formulae,[76, 79] correction for error (both systematic and random) when estimating the weight distribution of the population with missing data should be feasible by using simple Bayesian methods.[157] Information on weight distributions of the in- and ex-utero portions of the population can therefore be combined to simulate a cohort with the weights of all fetuses at risk at the beginning of each gestational week.

4.8 Conclusions

The missing intrauterine weight data in conventional birth weight references have resulted in a case definition for SGA that reflects ‘‘a birth weight below the population 10th percentile, corrected for gestational age’’ (p. 870)[22] only at term ages. At preterm ages, the threshold for SGA reflects a much lower percentage of the total at-risk population, leading to a case definition of SGA that is inconsistent across gestational ages. This case definition is problematic for epidemiologic studies, where exposures of interest are often associated with gestational duration and therefore can affect case status through mechanisms independent of their effect on weight. To correct the missing-data bias that currently exists in studies of fetal growth restriction, the following changes are needed:

1. References to assess neonatal weight must be developed that reflect the weight distributions of all fetuses in the population at the beginning of a given gestational week, not just livebirths. By definition, preterm births are not ‘‘normal’’ pregnancies and should therefore not be used to characterize the growth patterns of the full conception cohort. Thus, the correct reference chart is neither a livebirth weight reference nor an intrauterine estimated fetal weight reference, but a perinatal one that combines the weights of both livebirths and fetuses in utero at each week of gestation. At preterm ages, it will constitute predominantly in utero weights; as term ages approach, livebirth weights will make up a larger and larger portion of the distribution.

> 2. The case definition of SGA must be established as the bottom 10 percent of the total population at risk of being small, not just those who happen to be born at a given week of gestation. To establish that an infant of 32 weeks is small for its gestational age, its size needs to be compared with that of all other pregnancies that progressed to 32 weeks, regardless of whether those pregnancies went on to end at 32 weeks or 40 weeks. Researchers should stop classifying as normal the weights of growth restricted preterm infants simply because there are many other growth-restricted preterm livebirths who are even smaller than they are. This case definition is particularly important for etiologic studies of growth restriction, to prevent differential misclassification of SGA cases as noncases. Until an unbiased reference is available, the use of birth-weight-for-gestational-age charts should be restricted to term ages.[39]

Adopting the same approaches to the missing data in neonatal weight charts as we would for missing data in other areas of epidemiology will likely do much to further our understanding of perinatal population health.

4.9 Supplemental material for manuscript 1: measurement error in EFW

In the preceding manuscript, the theoretical bias resulting from the use of the conventional classification of "small-for-gestational-age" in the study of fetal growth was outlined. The extent of the bias was quantified through a simple simulation, and, based on these results, it was concluded that the bias associated with conventional classification of SGA was large enough to be of substantive importance. The quantification of the magnitude of the bias, however, relied on estimates of fetal weight derived from obstetrical ultrasound measurements. Since ultrasonographic estimates of fetal weight are known to have measurement error, further consideration of the potential impact of this error on our conclusions is important. In this thesis section, the impact of error in estimates of fetal weight will be considered. Through additional calculations, we will demonstrate that the previously demonstrated bias arising from the use of the conventional classification of SGA is unlikely to be solely attributable to measurement error.

Estimated fetal weight formulae

Evidence that preterm births are smaller than their in utero peers is derived primarily from studies that have compared estimated fetal weight distributions to the weight distributions of preterm births.[17-19] It could therefore be argued that if EFW formulae overestimate fetal weight, the apparent difference between the weight distributions of the intrauterine and livebirth populations could be largely the result of systematic measurement error. In a study by Hediger, [65] however, the existence of a genuine difference in mean weights between the two populations was demonstrated. A 32-week ultrasound was performed on the *in utero* study population, and the 32- week EFWs of fetuses that were subsequently born preterm were compared to the 32-week EFWs of fetuses that went on to term births. Fetuses that were subsequently born at preterm ages were shown to be already significantly smaller (121 grams, p-value= 0.009) than their peers that were not born preterm, suggesting a genuine difference in growth between preterm and term infants.

Since these results were based on the weights of only 46 preterm births, however, and the research population in this study may not have been representative of a general obstetric population, we repeated the analysis with data from the Royal Victoria Hospital to confirm Hediger's findings. With our cohort of n=3015 pregnancies delivered between 2001-2004,

including 161 preterm births, we were able to reproduce these results. The 32-week estimated fetal weights of fetuses that were subsequently born preterm were 79 grams smaller than the 32-week estimated fetal weights of fetuses that were subsequently born at term. In relation to the average fetal size at 32 weeks, this difference would be comparable to a 140 gram difference among 40 week old infants in the Canadian population,[12] making the difference both clinically and statistically significant ($p<0.0001$). These results support the conclusion that preterm births are, on average, smaller than their *in utero* peers of similar gestational age, and the discrepancy between the intrauterine and birth weight distributions is not the result of systematic overestimation of weight by EFW formulae.

The difference between the *mean* weights of preterm births and their *in utero* peers is likely not solely an artefact of measurement error, but the impact of measurement error is still important to consider because of its impact on the variance of the weight distributions. The total error in EFW formulae is estimated to be in the range of 7-15%, of which the majority is believed to be random error.[85] This error in estimated fetal weight is problematic for clinical practice, where an accurate estimate for each fetus is needed to guide decision-making, but is less important for the creation of population distributions and reference values. In the presence of error that is predominantly random, a distribution's mean value will be correct, but the spread of the distribution will be broadened. Given that the intrauterine weight distribution is shifted to systematically higher weights than the birth weight distribution at preterm ages, the broadening of the fetal weight distribution due to random error would be expected to *reduce* the discrepancy between the two distributions, not *create* the bias.

A final concern with most EFW formulae comes from their use of a regression-derived formula to estimate fetal weights.[79-83] Since the values predicted from a regression model will have a more narrow distribution than that of the original values, the distribution of weights produced by most EFW regression formulae may be artificially narrow. If so, the value of the 10^{th} percentile of the intrauterine population would be at an artificially higher value, which would in turn overestimate the discrepancy between the SGA thresholds of the intrauterine and livebirth populations. Although this concern is true with the majority of EFW formulae, it is not a major concern for the formula used to estimate fetal weight in the Norwegian standard used in our simulation. Instead of being based on a regression formula, the estimated fetal weights in the

Norwegian standard were calculated through a volumetric approach to the estimation of weight. In this formula, biometric measurements are combined to estimate the volume of the head and trunk, as shown in figure 4-2. This approach makes more physiologic sense than regression-derived approaches, and has been shown to have lower random and systematic error than other formula at the extremes of the birth weight continuum.[84]

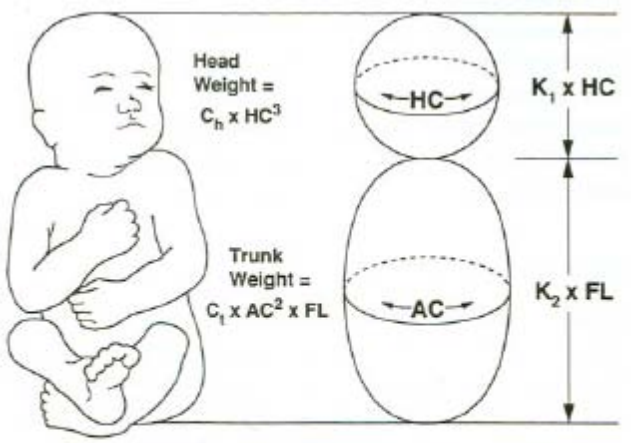


Figure 1. Two-compartment model of fetal weight. AC = abdominal circumference; FL = femur length; HC = head circumference; K_2, C_1 = constants. Algebraically, $C_1 = DK_2/6\pi$ and $C_h = DK_1/6\pi$, where D = density.

Figure 4-2. Derivation of Combs' formula to estimate fetal weight.[84]
Reproduced with permission, copyright owner Lippincott, Williams, & Wilkins.

Even with distributions of EFW derived from regression-based formulae such as Hadlock's,[79] the narrowing of the distribution is unlikely to play a major role. The magnitude of the error in EFW may be large enough to impede clinical decision making, but the correlation between estimated fetal weight and actual birth weight is still very high. The estimated weight produced by Hadlock,[79] for example, was shown to explain 96.5% of the variance in birth weight. With a large amount of variance explained by the regression parameters such as this, the distribution of predicted weights not likely be meaningfully more narrow than the actual distribution of weights in the population. To illustrate this point, a population of 100,000 infants with a similar weight distribution (mean and standard deviation) as that of 40-week old male births in the Canadian population[12] was simulated. An independent variable explaining 96.5% of the variability in weights in the population was built, as well as the weights predicted by a regression model regressing birthweight on this new explanatory variable. A plot of the two distributions (Fig 4-3) shows that the distribution of predicted weights is virtually identical to that of the simulated birth weights, and that the 10th percentile (SGA) threshold of the predicted weights is only 10

grams higher than that of the simulated birth weight distribution. Even if the amount of variability in birth weight explained by an EFW formula was as low as 85%, the impact on the distribution of predicted weights would still be minimal, with a difference in the 10th percentile thresholds of less than 45 grams. (Fig 4-4). One could speculate that this minor amount of narrowing of the estimated fetal weight distribution may be offset by the broadening of the distribution as a result of random error.

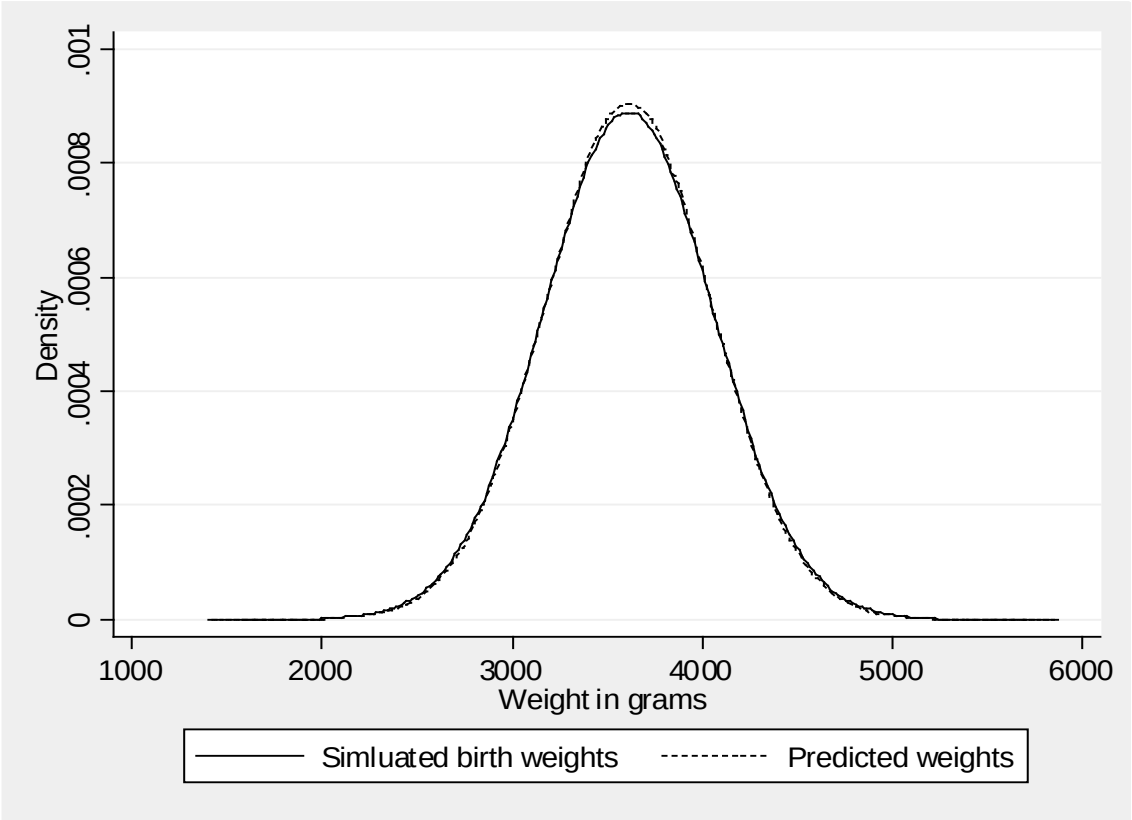


Figure 4-3. Distribution of simulated birth weights following the distribution of 40-week male fetuses in the Canadian population[12] and distribution of weights predicted by a regression model explaining 96.5% of the variability in birth weights.

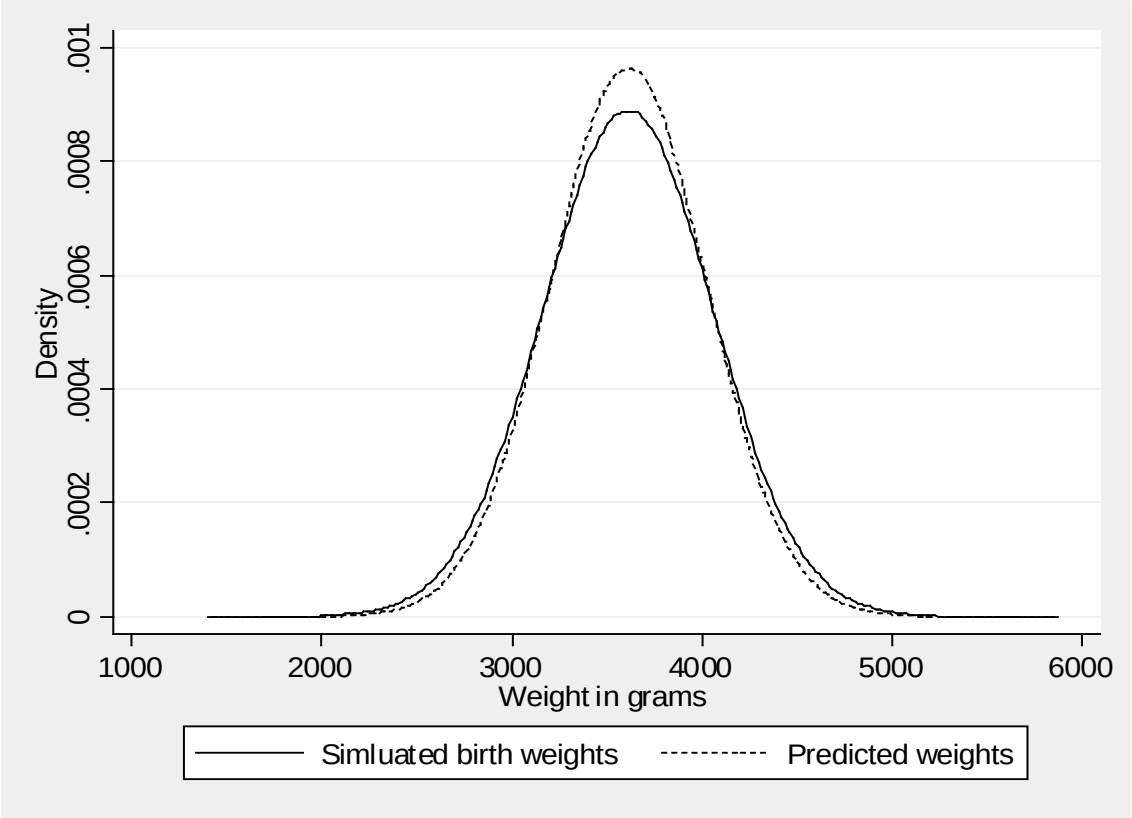


Figure 4-4. Distribution of simulated birth weights following the distribution of 40-week male fetuses in the Canadian population[12] and distribution of weights predicted by a regression model explaining 85% of the variability in birth weights.

Estimated fetal weight standards

In addition to concerns with the formulae used to estimate fetal weights, issues related to the construction of the intrauterine standards also need to be considered. Most intrauterine standards are based on relatively small samples that may not be representative of general obstetrical populations (The standard of Hadlock[91] was based on 392 white, middle class women from the United States mid-west region, the standard of Marsal [16] for Swedish populations was based on 86 pregnancies, while the Norwegian standard used in the previous manuscript was based on 635 pregnancies). If pregnancies selected for the construction of the standards were larger in size (perhaps fewer fetuses of smokers or pregnancies complicated by pre-eclampsia, both of which would be associated with smaller infants), the discrepancy between the conventional birth weight reference and the intrauterine standard would be exaggerated, and the extent of the bias that we reported would be inflated. We therefore repeated our calculation of the percent of infants misclassified as SGA by the use of a conventional standard (as presented in Table 4-1) using data from the Royal Victoria Hospital. The large, unselected sample

in this cohort overcomes concerns with selection bias possibly present in the smaller, research-based populations used to create existing reference charts. Since estimates of fetal weight were only available for 32 weeks, we assumed that the percent misclassification decreased at a constant linear rate until term (37 weeks), when the percent misclassification of SGA was assumed to be zero. As a conservative estimate, the percent misclassification prior to 32 weeks was held constant at the same percentage as that of 32-weeks, though in reality, it would likely increase with decreasing gestational age. The results of this sensitivity analysis are presented in Table 4-2. After correcting for the misclassification of infants in the smallest 10 percent of the total cohort that were classified as "appropriate-for-gestational- age" by the conventional birth weight reference, the original relative risk of 2.72 was re-calculated to a relative risk of 3.52. Since the extent of the bias according to these calculation was slightly larger than that calculated using the Norwegian populations, this provided confirmatory evidence that our initial conclusions were valid, despite the potential non-representativeness and small samples sizes of existing intrauterine standards.

Error in the estimation of fetal weight poses an important challenge in the study of fetal growth, and ongoing studies to improve estimated fetal weight formulae and estimated fetal weight standards[158] should help better define the true extent of the discrepancy between the weights of preterm births and their *in utero* peers. Nevertheless, although our knowledge on the magnitude of the discrepancy between the two populations is crude, it can still be concluded that the magnitude is large enough to be of real substantive importance and introduce bias to epidemiologic studies of fetal growth restriction.

TABLE 4-2. Bias to the estimate of relative risk of SGA* introduced because of misclassification of preterm SGA infants in a simulated population of 10,000 exposed and 10,000 unexposed infants, with the percentage of misclassification based on Royal Victoria Hospital data

Gestational age at birth (weeks)	***% SGA infants misclassified as AGA****	***Unexposed cohort (gestational age 39.0±2.3 weeks[†])***			***Exposed cohort (gestational age 39.0±2.3 weeks[†])***		
		No. births	Observed No. SGA births[‡]	Corrected No. SGA births	No. births	Observed No. SGA births[§]	Corrected No. SGA births
25	70				1	0.3	0.9
26	70				1	0.3	0.9
27	70				11	3.0	10.0
28	70				17	4.6	15.4
29	70	1	0.1	0.3	66	18.0	59.8
30	70	1	0.1	0.3	134	36.4	121.5
31	70	7	0.7	2.3	225	61.2	204.0
32	70	29	2.9	9.7	358	97.4	324.6
33	56	89	8.9	20.2	602	163.7	372.1
34	42	274	27.4	47.2	793	215.7	371.9
35	28	534	53.4	74.2	1,099	298.9	415.2
36	14	999	99.9	116.2	1,246	338.9	394.1
37	0	1,341	134.1	134.1	1,188	323.1	323.1
38	0	1,723	172.3	172.3	1,159	315.2	315.2
39	0	1,688	168.8	168.8	995	270.6	270.6
40	0	1,413	141.3	141.3	776	211.0	211.1
41	0	964	96.4	96.4	558	151.8	151.8
42	0	558	55.8	55.8	350	95.2	95.2
43	0	238	23.8	23.8	211	57.4	57.4
Total		9859	985.9	1063	9790	1958	2731.5
Risk of SGA per 100			10	10.8		27.2	37.9
Relative risk of SGA[#]						2.72	3.52

*established by calculating the percent of infants below the 10th percentile of an unselected intrauterine population at the Royal Victoria Hospital in Montreal, Canada, that were not identified as SGA by a Canadian birth weight reference[12] AGA=appropriate-for-gestational age

†cohort distribution truncated at 43 weeks of gestation therefore total number of births does not add to 10 000
‡number of SGA births observed with use of a live birth weight-based reference assuming risk of SGA is 10 percent among unexposed
§number of SGA births observed with use of a live birth weight-based reference if observed relative risk among exposed is 2.72
compared to unexposed cohort

5. Potential alternatives: customized birth weight percentiles

5.1 Preamble to manuscript 2

In the previous chapter, the bias in conventional birthweight percentiles and thresholds for "small-for-gestational-age" was outlined. In the following chapters, the focus shifts from a critical appraisal of conventional charts to an exploration of possible alternatives. A "perinatal" weight reference that reflects the weights of both ongoing pregnancies and births at each gestational age was suggested in the previous chapter as a more methodologically appropriate approach to development of weight-for-age percentiles, but such a reference does not exist as of yet. Further, the relative merits of existing alternatives deserve consideration before determining the most appropriate approach for the classification of fetal growth. The next two chapters therefore evaluate several of the more promising alternatives to conventional charts, "customized" birthweight percentiles and "conditional" fetal growth percentiles.

In this chapter, "customized" birthweight percentiles are considered. Customized birthweight percentiles are weight percentiles that have been individualized to account for maternal influences on fetal growth, such as maternal height, parity, ethnicity, and pre-pregnancy BMI.[104] The idea behind them is an intuitive one: if a mother is a short, small woman, her infant would be expected to be small as well (since maternal characteristics have been shown to be significant predictors of birthweight), and the assessment of fetal growth should be able to take this into account. The calculation of customized birthweight percentiles is therefore, in essence, a formalization of clinical intuition on expected birthweights. Customized birthweight references were first proposed in the literature over 15 years ago,[104] and numerous studies have subsequently shown that customized percentiles are better than conventional birthweight charts at identifying infants at increased risk of adverse perinatal outcomes.[106, 111, 114] Customized percentiles have steadily gained acceptance in the clinical and scientific community,[116] and have been recommended for clinical use in national practice guidelines.[115]

Work done by our group,[1] however, suggested that the apparent benefits of customised birthweight percentiles may instead be largely an artefact of "confounding" by gestational age.

Infants identified as “SGA” by the customized standard were more likely to be born preterm, and preterm birth itself is known to be associated with increased risks of adverse perinatal outcomes. The study demonstrated that the high relative risks of adverse outcome among infants classified as “SGA” by the customized standard were greatly attenuated when gestational age was included as a covariate in logistic regression models, suggesting that it was differences in gestational age, rather the process of “customizing” for maternal characteristics, that were important in identifying high-risk infants. This result was interesting, because in addition to adjusting for maternal characteristics, at preterm ages customized percentiles are also based on the weight distribution of the intrauterine population rather than the weight distribution of births. The reduction of relative risk seen when adjusting for gestational age of customized “SGA” infants could therefore be interpreted as assessing the impact of using an intrauterine-based standard, rather than a birthweight-based standard, at early gestational ages. This manuscript, “Customised birthweight percentiles: does adjusting for maternal characteristics matter?”(BJOG, 2008; 115:1397–1404), seeks to test this idea.

5.2. Title Page & Footnotes

Title: Customised birthweight percentiles: does adjusting for maternal characteristics matter?

Authors: Jennifer A HUTCHEON[1] BSc, Xun ZHANG[2] PhD, Sven CNATTINGIUS[3] MD PhD, Michael S KRAMER[1,2] MD, Robert W PLATT[1,2] PhD

Affiliations: McGill University Departments of Epidemiology & Biostatistics[1] and of Pediatrics[2], Montréal, Canada; and Karolinska Institute Department of Medical Epidemiology, Stockholm, Sweden[3]

Corresponding author:
Jennifer Hutcheon
The Montreal Children's Hospital Research Institute
4060 Rue Sainte Catherine Ouest #205
Westmount QC, Canada
H3Z 2Z3
Email: jennifer.hutcheon@mail.mcgill.ca
Tel: (514) 934-1934 x-23283
Home tel: (514) 288-9264
Fax: (514) 412-4331

Running title: Customised birthweight percentiles

KEY WORDS: Infant, Small for gestational age; Intrauterine growth retardation; Reference standards; Stillbirth; Infant mortality

Disclosure of interests: we have no conflicts of interest to declare

Funding: JAH is the recipient of a Doctoral Research Award from the Canadian Institutes of Health Research, RWP is a *Chercheur-Boursier* of the *Fonds de la recherche en santé du Québec*. The Montreal Children's Hospital Research Institute receives operating funds from the *Fonds de la recherche en santé du Québec*.

5.3 Abstract

Objective: To determine if the improved prediction of risk for perinatal mortality obtained with the use of a customised birthweight standard can also be obtained with the use of a non-customised, but intrauterine-based standard.

Design: Population-based cohort study.

Setting: Sweden.

Population: Births in the Swedish Medical Birth Register between 1992- 2001 (n=782,303) with complete data on birthweight, gestational age, sex, maternal age, pre-pregnancy body mass index, height, parity, and ethnicity.

Methods: We calculated the relative risks (RR) of stillbirth and early neonatal mortality among small-for-gestational-age (SGA) births as established by 1) a customised standard, 2) a population standard based on birthweights and 3) a population standard based on a best estimate of intrauterine weights.

Main outcome measures: stillbirth and early neonatal mortality (< 7 days).

Results: The RRs of stillbirth and early neonatal mortality among SGA births as classified by the intrauterine standard were similar to those among SGA births as classified by the customised standard, and much higher than those among SGA births as classified by the birthweight standard.

Conclusions: A non-customised, but intrauterine-based standard has a similar ability to predict risk for stillbirth and early neonatal mortality as a customised birthweight standard. The process of customising population weight-for-gestational-age standards to account for maternal characteristics does little to improve prediction of perinatal mortality.

5.4 Introduction

Customised birthweight percentiles, first proposed by Gardosi and colleagues,[104] are weight-for-gestational-age charts that have been individualised to account for maternal influences on fetal growth. By incorporating information generally believed to have a physiological influence on fetal growth (such as maternal height, pre-pregnancy body mass index (BMI), parity, ethnicity, and fetal sex), customised percentiles were designed to better differentiate between infants who are small because their *in utero* growth has been restricted and infants who are small but have reached their individual growth potential.[105] Customised birthweight percentiles have consistently been shown to be superior to conventional birthweight-for-gestational-age percentiles in predicting perinatal morbidity and mortality,[106, 111, 114] and as a result, have been recommended for clinical use by practice guidelines of the British Royal College of Obstetricians and Gynaecologists.[115]

Before customised birthweight charts replace conventional weight-for-age charts in clinical practice, however, a better understanding of their properties is needed. Although the reported benefits of customised birthweight percentiles are generally attributed to their adjustment for maternal characteristics, an alternative explanation for their improved ability to predict perinatal morbidity and mortality is possible. Customised birthweight percentiles have a second methodological difference from conventional birthweight-for-gestational-age charts, in addition to adjusting for maternal characteristics. At earlier gestational ages, the customised percentiles are based on Hadlock's proportionality formula,[105] a formula in which an infant's predicted "optimal" birthweight is expressed as a proportion of its "optimal" weight predicted for 280 days (40 weeks) according to the intrauterine growth curve of Hadlock.[91] In essence, the normative values in customised percentiles at younger gestational ages are based on the distribution of the best estimate of intrauterine weights, whereas conventional birthweight charts are based on the weights of live births. The reported benefits of customized percentiles could therefore be attributed to either 1) the process of adjusting for maternal characteristics or 2) the incorporation of an intrauterine standard instead of a birthweight standard at younger gestational ages. Before concluding that the process of customising weight-for-gestational-age percentiles for maternal characteristics is beneficial, the separate contributions of each of these two methodological differences need to be understood.

Based on work previously done by our group,[1] we had reason to believe that the overall improved prediction of mortality obtained with the use of customised birth weight charts may be due to methodological differences between intrauterine and conventional charts, rather than the process of adjusting for maternal characteristics. We therefore hypothesised that the improved prediction of perinatal morbidity and mortality previously demonstrated with the use of customised birthweight percentiles could also be obtained with a non-customised, but intrauterine-based standard, and that the regression-based adjustment for maternal characteristics may be an unnecessary step.

In this study, our first objective was to assess whether the improved prediction of perinatal mortality obtained through the use of a customised birthweight standard can also be obtained through the use of a non-customised, but intrauterine-based standard. Our second objective was to quantify the extent to which the maternal characteristics in the customisation model are able to explain variability in birthweight, in order to understand the amount of additional information that these variables provide to the prediction of optimal fetal weight.

5.5 Methods

Study population:

The study population was drawn from singleton births ≥ 28 weeks of gestation in the Swedish Medical Birth Register between the years 1992 to 2001. The register contains information on 98-99 percent of births in Sweden, including stillbirths from 28 weeks of gestation.[159, 160] The accuracy of the gestational ages, birthweights, and stillbirths in the register has previously been validated.[159, 160] We excluded infants with congenital anomalies and infants with missing data on sex, birthweight, gestational age, or maternal covariates (height, pre-pregnancy weight, parity, age, or place of birth), leaving 81.5% of the original population. Further details on the Swedish Medical Birth Register and the final study sample are provided elsewhere.[1, 106] The study was approved by the McGill Faculty of Medicine Institutional Review Board.

Calculation of customised and population percentiles:

Customised birthweight percentiles were calculated according to previously published methods.[104, 105] Briefly, an "optimal" birthweight for 280 days of gestation was calculated based on covariates obtained from stepwise multiple regression (maternal height, pre-pregnancy BMI, ethnicity, parity, and fetal sex), then this weight was extrapolated to the optimal weight for the gestational age at birth using Hadlock's proportionality formula. This formula expresses "optimal" predicted birthweight at earlier gestational ages as a proportion of the predicted weight at 280 days, using the fetal growth curve of Hadlock[91] to determine the trajectory through which the 280-day weight is reached. Customised percentiles were then calculated based on the discrepancy between optimal and actual birthweight.

Population-based percentiles were calculated in two alternative ways: 1) a sex- and gestational week-specific birthweight reference, based on the weight distribution of live births in this population[1] and 2) the intrauterine (ultrasound) estimated fetal weight-for-gestational-age percentile chart published by Hadlock.[91] Although an intrauterine standard is available for Swedish populations,[16] Hadlock's standard was chosen for reasons of comparability, since the proportionality formula in the commonly used and publicly available customised standard is derived from Hadlock's standard.[105] Small for gestational age (SGA), was defined as a birthweight

below the 10th percentile, based on each of the birthweight, intrauterine weight, and customised standards [denoted as: SGA(birthweight), SGA(intrauterine) and SGA(customised), respectively].

Outcomes:

The primary outcomes in this study were the occurrence of stillbirth and early neonatal death. Stillbirth was defined as a fetal death at 28 weeks of gestation or later, and included both antepartum and intrapartum fetal deaths. Early neonatal mortality was defined as the death of a live-born infant before 7 days of age.

Statistical analyses:

The relative risks of stillbirth among SGA infants as defined by each of the 3 standards were calculated with 95% confidence intervals (CI), using the infants classified as 'non-SGA' by the same standard as the referent group. Denominators in the calculations of risk were based on the number of ongoing pregnancies at risk of stillbirth at each gestational age. The number of fetuses at risk (rather than live births) is the methodologically appropriate denominator in the study of stillbirths, since all unborn fetuses are at risk for stillbirth, and by definition, live births are no longer at risk of being stillborn.[153, 161] Because Hadlock's intrauterine standard is not sex-specific,[91] a generalized linear model (binomial family, log link) was used to calculate sex-adjusted relative risks for the intrauterine standard. Relative risks were calculated separately for term/post-term (≥37 weeks), mild preterm (34-36 weeks) and moderate-severe preterm (28-33 weeks) periods. Calculations were repeated using the outcome of early neonatal death. Fetuses at risk were also used as the denominator in the calculation of risk for early neonatal death[153] because all live fetuses are at risk of birth, and hence of neonatal death. Moreover, the risk factor of interest, intrauterine growth restriction, is known to lead to preterm birth (due to either obstetrical intervention or, to a lesser degree, spontaneous birth).[64] Since preterm birth is a downstream effect of the risk factor, stratifying on birth status (by restricting to live births) would inappropriately adjust out the effects of poor fetal growth on mortality.

Multivariable linear regression was used to quantify the extent to which variables in the customisation model were able to explain variability in term birthweight (259-293 days, inclusive, similar to previously published models[104, 105]), as established by the adjusted R^2. Since we were interested in understanding the contribution of maternal characteristics independent

of the contributions of sex and gestational age (which are already "customised" for in age- and sex-specific weight-for-gestational-age population standards), two separate models were built. The first contained only sex and gestational age at birth (linear and quadratic terms) as independent predictors, while the second additionally included the maternal characteristics (height, pre-pregnancy BMI, ethnicity, and parity). In both models, gestational age was centred at 280 days. The amount of variability explained by maternal characteristics alone was established as the incremental difference between the model with only sex and gestational age and the "full" customisation model. Statistical analyses were conducted using Intercooled STATA 9.0 (Stata Corporation, College Station, TX).

5.6 Results

Descriptive characteristics of the 782,303 infants and mothers in the final study population are presented in Table 5-1. The percentage of infants classified as SGA by each of the standards is shown in Table 5-2 according to gestational age at birth. While the percentage of infants classified as SGA by the population-based birthweight standard remained close to 10 percent across all gestational ages, the percentage of infants classified as SGA by the customised standard was much higher at early gestational ages, as high as 35% at 28-33 weeks. The percentages of SGA obtained with the intrauterine standard were very similar to those produced by the customised standard: 34% of infants born at 28-33 weeks were classified as SGA.

When the relative risks of stillbirth were calculated for SGA infants classified by each of the three standards (Table 5-3), two trends were apparent. First, the relative risks of stillbirth among SGA(customised) and SGA(intrauterine) were extremely similar overall and within each gestational age group, supporting our study hypothesis. Second, the difference between the RR(customised) or RR(intrauterine) and the RR(birthweight) varied by gestational age. At term ages, the relative risks produced by the different standards were fairly similar, but at preterm ages, the relative risks obtained from the intrauterine and customised standards were significantly higher than those obtained from the birthweight standard. The overall improved prediction of stillbirth obtained with the customised and intrauterine standards (RR=6.1 for customised standard, RR=6.2 for intrauterine standard) compared to the birthweight standard (RR= 3.8) was therefore driven primarily by improved classification of SGA at early gestational ages. The 10^{th} percentile threshold at 30 weeks as established by each of the three standards is shown in Figure 5-1 in relation to the weights of stillbirths at this age. The 10^{th} percentile of the birth weight standard was nearly 300 grams lower than that of the intrauterine standard, while the 10^{th} percentiles produced by the customised standard (shown as a distribution rug plot below the histogram, since each infant's customised 10^{th} percentile weight is different), were clustered around the value of the intrauterine 10^{th} percentile.

Table 5-4 summarizes the corresponding results for early neonatal mortality. As with stillbirth, relative risks of early neonatal mortality were similar for infants classified as SGA by the

intrauterine standard and infants classified as SGA by the customised standard. Both were higher than the relative risk obtained with the conventional birthweight-for-gestational-age chart's classification of SGA. The overall improved prediction of early neonatal mortality obtained with the customised and intrauterine standards was again primarily derived from improved prediction at preterm gestational ages, with the three standards yielding comparable relative risks at term.

Multivariable linear regression was used to compare explained variance in term birthweight (Table 5-5) between models with and without customisation for maternal characteristics. The regression model that included only gestational age and sex (Model 1) explained 17 percent of the variance in term birthweight. Once the sex and gestational age of the infant were known, information on maternal characteristics resulted in a modest improvement in the prediction of birthweight. Including maternal characteristics into the model explained an additional 7% of variance in birthweight (24% explained by the full model vs 17% explained by gestational age and sex alone).

Table 5-1. Descriptive characteristics of the study population of n=782 303 births in the Swedish Medical Birth Register, 1992 to 2001.

Maternal characteristic	**mean ± SD or n (%)**
Pre-pregnancy BMI* (kg/m^2)	23.9 ± 4.0
Maternal height (cm)	166.3 ± 6.2
Maternal age (years)	28.9 ± 5.0
Parity (%nulliparous)	325,247 (41.6)
Country of birth (% Nordic[†])	680,960 (87.1)
Fetus/infant characteristic	
Birth weight (grams)	3,566.1 ± 552.2
Gestational age at birth (weeks)	39.4 ± 1.7
Stillbirth	2,354 (0.3)
Early neonatal death (< 7 days)	815 (0.1)

*BMI=Body Mass Index

[†]Nordic= Sweden, Norway, Denmark, Finland, Iceland

Table 5-2. Percentage of infants identified as small-for-gestational age by the customized, birthweight, and intrauterine standards in n=782,303 births in the Swedish Medical Birth Register, 1992-2001.

	Birth weight SGA n (%)	**Intrauterine SGA n (%)**	**Customised SGA n (%)**
Delivery 28-33 weeks (n= 8,116)	801 (9.9)	2,725 (33.6)	2,840 (35.0)
Delivery 34-36 weeks (n= 28,472)	2,821 (9.9)	4,406 (15.5)	4,595 (16.1)
Delivery ≥37 weeks (n= 745,715)	73,613 (9.9)	65,966 (8.9)	70,854 (9.5)
All ages (n=782, 303)	77,235 (9.9)	73,097 (9.3)	78,289 (10.0)

Table 5-3. Relative risk of stillbirth among small-for-gestational age (SGA) infants, as established through customized, birthweight, and intrauterine standards in n=782 303 births in the Swedish Medical Birth Register, 1992-2001.

Gestational age	**Standard**	**SGA (<10th percentile)**		**Non-SGA***		**Relative risk† [95% CI]**
		At-risk population n	**Stillbirths n (risk per 1000)**	**At-risk population n**	**Stillbirths n (risk per 1000)**	
All ages	Birthweight	77,235	687 (8.9)	705,068	1,667 (2.4)	**3.8 [3.4, 4.1]**
	Intrauterine	73,097	906 (12.4)	709,206	1,448 (2.0)	**6.2 [5.7, 6.7]**
	Customized	78,289	952 (12.2)	704,014	1,402 (2.0)	**6.1 [5.6, 6.7]**
Delivery 28-33 weeks	Birthweight	77,235	144 (1.9)	705,068	469 (0.7)	**2.8 [2.3, 3.4]**
	Intrauterine	73,097	337 (4.6)	709,206	276 (0.4)	**12.2 [10.4, 14.3]**
	Customized	78,289	342 (4.4)	704,014	271 (0.4)	**11.4 [9.7, 13.3]**
Delivery 34-36 weeks	Birthweight	76,434	152 (2.0)	697,753	310 (0.4)	**4.5 [3.7, 5.4]**
	Intrauterine	70,372	200 (2.8)	703,815	262 (0.4)	**7.9 [6.6, 9.5]**
	Customized	75,449	212 (2.8)	698,738	250 (0.4)	**7.9 [6.5, 9.4]**
Delivery ≥37 weeks	Birthweight	73,613	391 (5.3)	672,102	888 (1.3)	**4.0 [3.6, 4.5]**
	Intrauterine	65,966	369 (5.6)	679,749	910 (1.3)	**4.2 [3.7, 4.7]**
	Customized	70,854	398 (5.6)	674,861	881 (1.3)	**4.3 [3.8, 4.8]**

* Reference category

† RR(intrauterine) adjusted for fetal sex

Table 5-4. Relative risk of early neonatal death among small-for-gestational age (SGA) infants, as established through customized, birthweight, and intrauterine standards in n=782 303 births in the Swedish Medical Birth Register, 1992-2001.

Gestational age	**Standard**	**SGA (<10th percentile)**		**Non-SGA***		**Relative risk† [95% CI]**
		At-risk population n	**Early neonatal deaths n (risk per 1000)**	**At-risk population n**	**Early neonatal deaths n (risk per 1000)**	
All ages	Birthweight	77,235	229 (3.0)	705,068	586 (0.8)	**3.6 [3.1, 4.2]**
	Intrauterine	73,097	297 (4.1)	709,206	518 (0.7)	**5.9 [5.1, 6.8]**
	Customized	78,289	328 (4.2)	704,014	487 (0.7)	**6.1 [5.3, 7.0]**
Delivery 28-33 weeks	Birthweight	77,235	43 (0.6)	705,068	182 (0.3)	**2.2 [1.5, 3.0]**
	Intrauterine	73,097	100 (1.4)	709,206	125 (0.2)	**8.3 [6.3, 10.8]**
	Customized	78,289	108 (1.4)	704,014	117 (0.2)	**8.3 [6.4, 10.8]**
Delivery 34-36 weeks	Birthweight	76,434	59 (0.8)	697,753	127 (0.2)	**4.2 [3.1, 5.8]**
	Intrauterine	70,372	71 (1.0)	703,815	115 (0.2)	**6.5 [4.9, 8.8]**
	Customized	75,449	80 (1.1)	698,738	106 (0.2)	**7.0 [5.2, 9.3]**
Delivery ≥37 weeks	Birthweight	73,613	127 (1.7)	672,102	277 (0.4)	**4.2 [3.4, 5.2]**
	Intrauterine	65,966	126 (1.9)	679,749	278 (0.4)	**4.9 [4.0, 6.1]**
	Customized	70,854	140 (2.0)	674,861	264 (0.4)	**5.1 [4.1, 6.2]**

* Reference category

† RR(intrauterine) adjusted for fetal sex

Table 5-5. Explained variance in multiple linear regression models of birthweight among n=688,529 term births (259-293 days gestation, inclusive) in the Swedish Medical Birth Register 1992-2001.

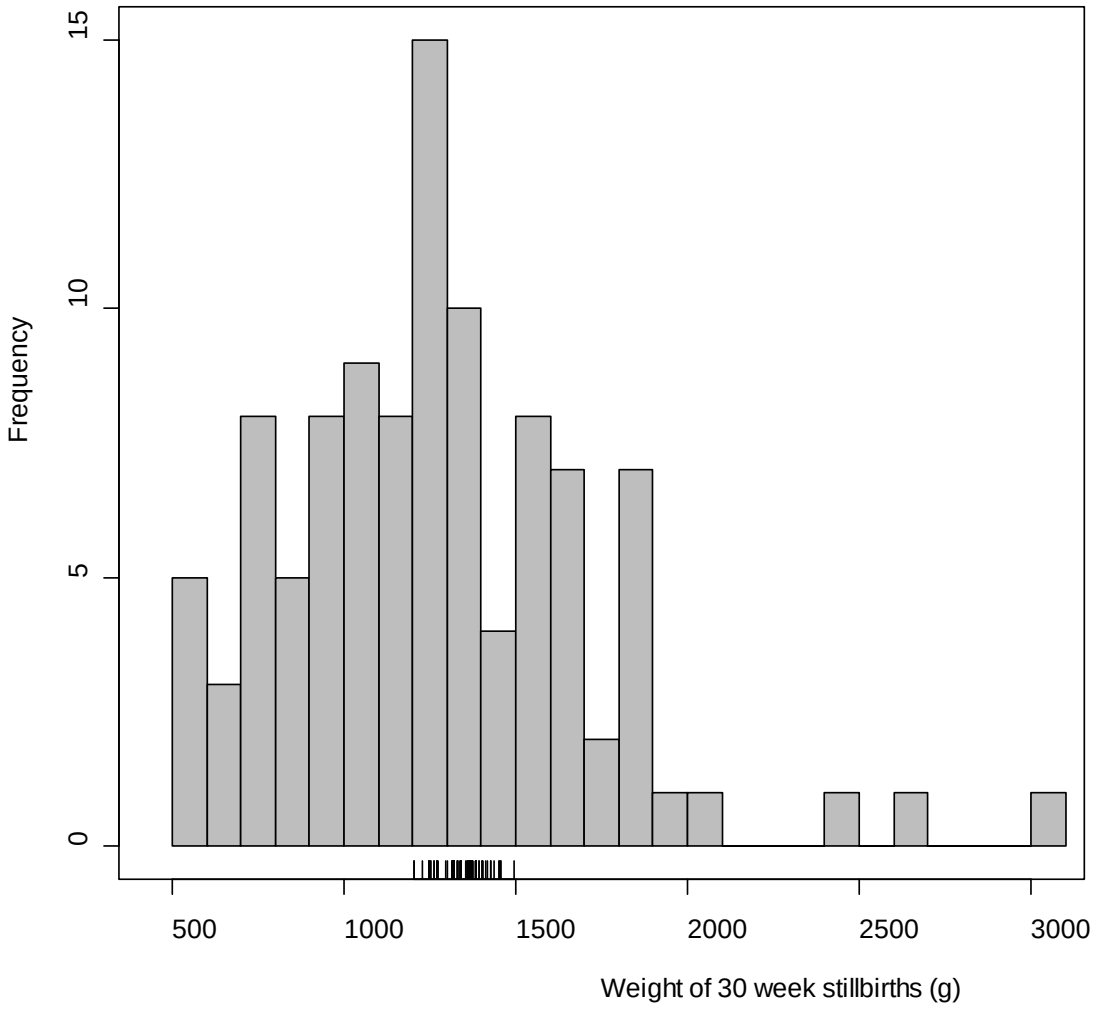
	Model 1 (gestational age and sex only)		**Model 2 (full customization model)**	
	Coefficient	[95% CI]	Coefficient	[95% CI]
Sex (male)	121.1	[119.0, 123.2]	120.6	[118.6, 122.6]
Gestational age* (days)	23.1	[22.9, 23.2]	22.2	[22.1, 22.4]
Gestational age^2 (days)	-0.2	[-0.2, -0.2]	-0.2	[-0.2, -0.2]
Parity†				
Para 1			146.7	[144.4, 148.9]
Para ≥ 2			177.5	[174.9, 180.2]
Pre- pregnancy BMI‡				
<18.5 kg/m^2			-153.2	[-159.1, -147.3]
25-29.9kg/m^2			107.1	[104.7, 109.6]
≥30kg/m^2			175.9	[172.1, 179.6]
Height§				
<160cm			-119.2	[-122.3, -116.0]
>170cm			120.7	[118.5, 123.0]
Ethnicity (Non-nordic)			-62.7	[-65.8, -59.7]
Intercept	3557.2	[3555.4, 3558.9]	3415.0	[3412.7, 3417.4]
Adjusted-R^2	**0.17**		**0.24**	

*Gestational age centred at 280 days

†Reference category nullipara

‡Reference category 18.5-24.9 kg/m^2

§Reference category 160-170 cm


Frequency
0
5
10
15
500
1000
1500
2000
2500
3000
Weight of 30 week stillbirths (g)

5.7 Discussion

In this study, we have shown that a non-customised intrauterine weight standard has a similar ability to predict perinatal mortality as a customised birthweight standard. We conclude that the improved prediction of perinatal morbidity and mortality by customised birthweight percentiles compared to conventional birthweight percentiles is not derived from their adjustment for maternal characteristics (which is widely believed to be responsible for their apparent benefits[116]) but rather, is derived from their use of Hadlock's intrauterine-based proportionality formula at preterm gestational ages. Maternal characteristics contributed little additional information to the customisation model compared to the information obtained from sex and gestational age at birth, explaining why customisation of population-based weight percentiles did little to further improve the prediction of mortality.

The validity of these conclusions is supported by the consistency of our results with previous reports on customised percentiles. The percentages of infants classified as SGA by the customised standard in this study were very similar to those recently reported by Groom and colleagues[162] (29.1% <34 weeks, 18.0% at 34-36[6] weeks, and 9.5% at ≥ 37 weeks), as well as to previous studies estimating the percentage of preterm live births classified as SGA by an intrauterine standard.[16, 99-101] The increased relative risk of perinatal mortality among our study subjects classified as SGA by the customised standard was similar to results of previous studies of customised percentiles.[106, 114] Unlike these earlier studies, however, we extended our evaluation of customised percentiles to additionally include a comparison with an intrauterine standard, and found no statistically significant difference in the relative risk of stillbirth or early neonatal death between the customised and intrauterine standard. Conventional birthweight charts are biased at preterm gestational ages because of the association between fetal growth restriction and preterm birth,[16-19, 96, 102] and our results demonstrate that correction of this bias through the use of an intrauterine standard[101, 163] improves identification of high-risk fetuses. Once this bias has been corrected, the process of customising for maternal characteristics does little to improve prediction of perinatal mortality. In the single customisation study that used a population standard based on intrauterine weights rather than birthweights, customisation was also found to have no benefits in the identification of growth restricted stillbirths.[109] The large, population-based sample in our study provided sufficient statistical precision to detect

differences and similarities between the customised, intrauterine, and birthweight standards in the prediction of perinatal mortality.

Although the proportion of variance in birthweight explained by our full customisation model was comparable to published values (e.g. R^2 =0.27 in a customisation model for a French population[114]), the majority of the variance explained was due to contributions of gestational age and sex. The absolute amount of variance explained by maternal characteristics was small (7%), which likely explains why the process of adjusting for maternal characteristics does not add meaningful information to distinguish truly growth-restricted infants from "small-but-healthy" infants. The minor effect of adjusting for maternal characteristics is consistent with the results of our recent publication based on the same Swedish study sample.[1] We found that controlling for the increased percentage of preterm infants classified as SGA by the customised standard led to a large reduction in the relative odds of stillbirth (Unadjusted OR=7.8, OR adjusted for gestational age =2.4). If adjustments for maternal characteristics were important in improving identification of stillbirth, such a large reduction would *not* have been seen. The small absolute amount of additional information provided by the maternal characteristics was a substantial proportion of the variance explained by the full customization model, however. This result serves to highlight that a much better understanding of the physiological influences on fetal growth is likely needed before individualized risk prediction can be successful. Although we adjusted for similar maternal characteristics as those included in previous customization models, it is possible that future customization models with additional independent predictors may lead to improved prediction of birth weight.

The relative risks presented in this paper should be interpreted with the limitations of the data in mind. As in previous studies, the gestational age of stillbirths was the age of birth, not the age of death, and it is unknown how much time elapsed between fetal demise and delivery. Likewise, the recorded weight of stillbirths was the weight at birth, not at the time of death. Any loss of weight between the time of death and time of delivery would be expected to overestimate the predictive value of weight-for-age charts.[164] Although these shortcomings may have introduced error to our estimates of risk, there is no reason to believe that such an error would have been differential among the three standards (since all calculations used the same data on weights and gestational ages). It is therefore unlikely to affect the conclusions of our

study, which focused on a comparison between intrauterine and customised standards rather than absolute effect.

CONCLUSION

Customised birthweight percentiles have already been recommended for clinical use by the UK's Royal College of Obstetricians and Gynaecologists,[115] and a recent editorial has called for American obstetricians to adopt their use as well.[116] The results of our study demonstrate that while the customisation for maternal factors does not impede the identification of high-risk infants, the process also provides little additional predictive benefit. The important contribution of past work by Gardosi and colleagues[104] appears to be not so much their regression model to "customise" for maternal characteristics, but their recognition of the inappropriateness of using a birthweight-based standard at preterm ages.[163] Since data on maternal characteristics are often missing, even in high-quality databases such as the Swedish Medical Birth Register used in this study, a non-customised, but intrauterine-based standard may be the most parsimonious and practical standard for the prediction of perinatal mortality in clinical practice.

5.8 Supplemental material for manuscript 2: Why customization for maternal characteristics help identify fetal growth restriction?

Rationale

Earlier in this chapter, we demonstrated that an intrauterine standard has a similar ability to predict risk of perinatal mortality as a customized standard, and that maternal characteristics explain only a small amount of the variability in birth weight (as seen by the customization model's adjusted R^2). In this section, the link between these two observations will be explored in further detail. How does the extent to which a group of maternal characteristics is able to explain variability in birthweight affect the "optimal" weights predicted for each infant by the customization model, and in turn, the customized percentiles assigned? If the "optimal" weights predicted by existing customization models do not lead to customized percentiles that are meaningfully different than conventional birthweight percentiles, how well *would* a set of maternal characteristics need to be able to explain variability in birthweight in order for the process of customization to be worthwhile? Of particular interest is the impact of a model's adjusted R^2 on the optimal weights predicted at the lower end of the birthweight continuum, since a major goal of customization is to help distinguish infants that are physiologically small from infants who were small because they had failed to reach their growth potential.

Methods

To explore the extent to which the variability in birthweight explained by of a group of maternal characteristics influences the optimal weights predicted by a customization model, a simulated cohort of 100,000 infants was created. The characteristics of this cohort (means and variances for birthweight, gestational age at birth, maternal height and pre-pregnancy BMI, as well as proportions for sex, ethnicity, and parity) were based on the characteristics of term births in the Swedish Medical Birth Register population described in section 5.5 (shown in Table 5-6). The intercorrelations between these variables were also simulated to be similar to those observed in the Swedish Medical Birth Register population (Table 5-7). Variables were simulated to be multivariate normal, and then converted to binomial variables where necessary (sex, ethnicity, parity).

In addition to the maternal characteristics included in previous customization models, an additional hypothetical maternal characteristic explaining variability in birthweight was simulated. This hypothetical characteristic could represent an as-yet-undiscovered biomarker or gene that is predictive of physiological differences in birthweight, and its inclusion allowed us to explore the relationship between the adjusted R^2 of a customization model and the "optimal" weights predicted by the model. This hypothetical characteristic was simulated to follow a standard normal distribution, and was correlated with birthweight at varying correlations of r=0, 0.25, 0.5, 0.75. For simplicity, this hypothetical characteristic was simulated to be independent of the other maternal characteristics. Although the correlations used in these simulations were chosen arbitrarily, we were particularly interested in the correlation of 0.50, because one of the strongest known predictors of infant birthweight is the birthweight of a sibling, which has a correlation of roughly 0.50.[165] The inclusion of our hypothetical maternal characteristic at a correlation of 0.50 would therefore roughly correspond to the inclusion of "sibling birthweight" in a customization model.[166]

Table 5-6. Descriptive characteristics of n= 688,529 term births in the Swedish Medical Birth Register, 1992-2001.

Maternal characteristic	**Mean**	**Std. Dev.**
Pre-pregnancy BMI (kg/m^2)*	23.9	4.0
Maternal height (cm)	166.2	6.2
	N	**%**
Parity (%nulliparous)	278,154	40.40
Ethnicity (% Nordic†)	598,960	87.0
Fetal characteristic	**Mean**	**Std. Dev.**
Birth weight for sex and gestational age Z-score	-1.2×10^{-07}	1.0
Gestational age (days)	279.4	7.9
	N	**%**
Sex (% male)	348,995	50.7

* BMI= body mass index

† Nordic=Sweden, Norway, Denmark, Finland, Iceland

Table 5-7. Correlation matrix of fetal-maternal characteristics and birthweight Z-score obtained from 688,529 term births in the Swedish Medical Birth Register, 1992-2001.

	Birth weight Z-score	Multiparous	Pre-pregnancy BMI	Maternal Height	Non-nordic ethnicity
Z-score	1.0000				
Multiparous	0.1816	1.0000			
BMI*	0.1805	0.0939	1.0000		
Height	0.1927	-0.0288	-0.0621	1.0000	
Non-nordic†	-0.0837	0.0161	0.0031	-0.2630	1.0000

* BMI= body mass index

† Non-nordic=Any country other than Sweden, Norway, Denmark, Finland, or Iceland

Multivariable linear regression customization models were generated with maternal height, pre-pregnancy BMI, ethnicity, parity, and the new hypothetical maternal characteristic as independent variables and sex- and gestational-age-specific birthweight Z-score as the dependent variable. Birthweight Z-scores were calculated using an internal standard (similar to that used to calculate birthweight percentiles in the preceding manuscript). The use of birthweight Z-score as the dependent variable (instead of birthweight) allowed the effects of the maternal characteristics to be separated from the effects of sex and gestational age, which are already "customized" for in conventional sex- and gestational age- specific charts. Models were generated using varying correlations of the hypothetical maternal characteristics and birthweight Z-score to produce customization models with different adjusted R^2s.

With each model, "optimal" birthweight Z-scores were predicted for each infant, and the distribution of "optimal" birthweight Z-scores in the simulated cohort was examined graphically. The percentage of customized birthweight Z-scores that were meaningfully different than the population average (i.e. the "optimal" weight predicted by a conventional birthweight-for-gestational-age chart, which is the population's 50^{th} percentile), as well as the number of "optimal" weight predicted to be below the population's 10^{th} percentile, were calculated.

Results and Discussion

The effect of increasing the correlation between birthweight and the hypothetical maternal characteristic on the customization model's adjusted R^2 is shown in Table 5-8. Not surprisingly, as the correlation between birthweight and the hypothetical maternal characteristic increased, the customization model's adjusted R^2 also increased, from a minimum of 0.09 when the hypothetical maternal characteristic was set to be uncorrelated with birthweight (i.e. only maternal height, parity, ethnicity, and pre-pregnancy BMI were included as predictors) to a maximum of 0.66 when the correlation between the added hypothetical maternal characteristic and birthweight was set at 0.75.

The adjusted R^2 of these models are important, because they influence the "optimal" birthweight Z-scores predicted for each infant. In the absence of any information on maternal characteristics, the best estimate of an infant's "optimal" weight is that of the average weight in the population, the 50^{th} percentile. Thus, with a null model (an adjusted R^2 of 0), all infants in the

population would be assigned an "optimal" weight that was equal to the population 50th percentile weight. This approach would be no better than a conventional birthweight-for-gestational-age chart in identifying fetuses that had failed to reach their own individual growth potential. As more and more information on maternal characteristics is obtained, the distribution of "optimal" weights predicted by the customization will become broader. Rather than being predicted to have an "optimal" birthweight that is equal to the population 50th percentile, the infant of a short, thin, nulliparous woman would be predicted to have a lower "optimal" birthweight. In the extreme case of a model with an R^2 of 1 (i.e. the model was able to perfectly predict all infants' birthweights based on maternal characteristics), the variability in "optimal" weights predicted by the customization model would be similar to the variability in birthweights in the population, allowing deviations in fetal growth to be easily identified.

Figures 5-2 to 5-5 illustrate the effects of this decrease in adjusted R2 on the "optimal" birthweight Z-scores predicted by the customization model. In these graphs, the birthweight Z-scores predicted by each of the customization models is overlaid on the distribution of actual birthweight Z-scores in the simulated population. While the distribution of optimal birthweight Z-scores predicted by a customization model with an adjusted R^2 of 0.66 was reasonably similar to that of the actual Z-scores (Fig 5-2), the distribution of predicted birthweight Z-scores became more and more narrow as the adjusted R^2 of the model decreased. When the customization was able to explain only 9% of the variability in birthweight Z-score (Fig 5-5), the "optimal" birthweight Z-scores were all clustered closely around the population average birthweight Z-score of 0, with a much reduced variability compared to the distribution of actual birthweight Z-scores in the population.

If a customization model was unable to explain any of the variability in birthweight Z-scores (i.e., a null model with an adjusted R^2 of 0), the predicted Z-scores for each infant would all be equal to a Z-score of 0, since, in the absence of any other information, the best estimate of "optimal" birthweight Z-score for an infant would be that of the population average. Although the customization model of Figure 5-5 was able to explain some of the variability in birthweight Z-scores, this amount was only modestly larger than that of a null model. As a result, the "optimal" birthweight Z-scores predicted for each infant were only modestly different than those that would have been predicted by a null model. Since conventional birthweight-for-gestational age

charts are, in essence, a null customization model (by incorporating no information on maternal characteristics and simply using the population 50th percentile weight for sex and age as the optimal weight for each infant), customized percentiles from a poor customization model and conventional percentiles wind up being highly similar. As a result, the classifications of SGA obtained with conventional and customized are also very similar, as demonstrated in the results of section 5.6 earlier in this chapter.

The relationship between the adjusted R^2 of a customization model and the distribution of predicted "optimal" birthweight Z-scores was further quantified. First, the number of infants whose birthweight Z-Score predicted by the customization model was clinically equivalent to the birthweight Z-score predicted by the conventional approach (the 50th percentile, a Z-score of 0) was calculated. A Z-score of 0.22SD was chosen as the minimal amount to establish a clinically meaningful difference, which corresponds to a 100 gram difference among 40 week old births in this population. As shown in the third column of Table 5-8, when the customization model was able to explain 66% of the variability in birthweight Z-scores, only 21% of the predicted optimal Z-scores were clinically equivalent to the conventional approach, only slightly higher than the 17% expected if a customization model was able to perfectly predict variability in birthweight Z-scores. Thus, with this model the customized birthweight Z-score for many infants was closer to their physiologically optimal weight than the estimate obtained simply by using the population average. As the predictive ability of the customization model decreased, the percent of infants with a customized "optimal" weight that was clinically equivalent to the population average increased. This can be seen graphically in Figures 5-2 to 5-5, where the width of the histogram bars has been set to 0.22SD. Thus, the two bars of the histogram on either side of 0 represent the percent of infants within 0.22SD of the population average. With an adjusted R2 of 0.09, over half of the optimal birthweight Z-scores predicted by the customization model were clinically equivalent to the optimal weight predicted by a conventional birthweight-for-gestational-age chart. As seen in Fig 5-5, over 52% of infants had a predicted birthweight Z-score that was clinically equivalent to the population average, and over 85% of infants were within 0.44SD of the population (corresponding to roughly 200 grams). This clustering of optimal customized birthweight Z-scores within 200grams of the population leads to 80% reference limits, and as a result, SGA thresholds, that are also highly similar between the customized and conventional birthweight charts.

The prediction of optimal birthweight Z-score at the lower end of the birthweight continuum is of particular interest, since a major goal of customization is to identify infants that are "small-but-healthy". In order to do so, the process of customization must be able to successfully predict that these "small-but-healthy" infants will have an optimal birthweight that is lower than the population average. The final column of Table 5-8 shows the number of "small-but-healthy" infants that each model is able to predict (Z-scores below the 10^{th} percentile of the actual distribution, corresponding to a birthweight Z-score of -1.28SD). With a customization model that explained 66% of the variability in birthweight Z-scores, 6 percent of the population was identified as having an "optimal" weight that was in the "small-but-healthy" range. As the predictive ability of the model decreased, the percent of "small-but-healthy" infants predicted by the model also decreased. Of the simulated cohort of 100,000 infants, the customized weights predicted by the poorest model (explaining 9% of the variance in birthweight Z-scores) only predicted that 4 infants should have a birthweight Z-score below 1.28 SD of this population's true distribution. Thus, although a primary motivation for customizing birthweight percentiles is to distinguish "small-but-healthy" infants from those that are truly growth restricted, the optimal weights produced by a customization model with a poor predictive ability will not be able to correctly estimate the small optimal weight of most of these small-but-healthy infants. The customized weights predicted by a model with a poor predictive ability produces weights that are largely clinically indistinguishable from the weights predicted by a conventional birth weight for gestational age chart. As a result, it is not surprising that the process of adjusting for maternal characteristics through a customization model provides little additional information to distinguish between infants who are physiologically versus pathologically small.

Table 5-8. Weights predicted by customization models with varying adjusted R^2s in a simulated cohort of 100,000 births

Correlation of maternal characteristic with Z-score	Adjusted R^2	Predicted weights within 0.22SD of population average n(%)	Predicted weights below 1.28SD of population average n(%)
0.75	0.66	21129 (21)	5889 (5.9)
0.50	0.35	28914 (28)	1561 (1.6)
0.25	0.16	41685 (42)	66 (6.6)
0.00	0.09	52024 (52)	2 (0.2)

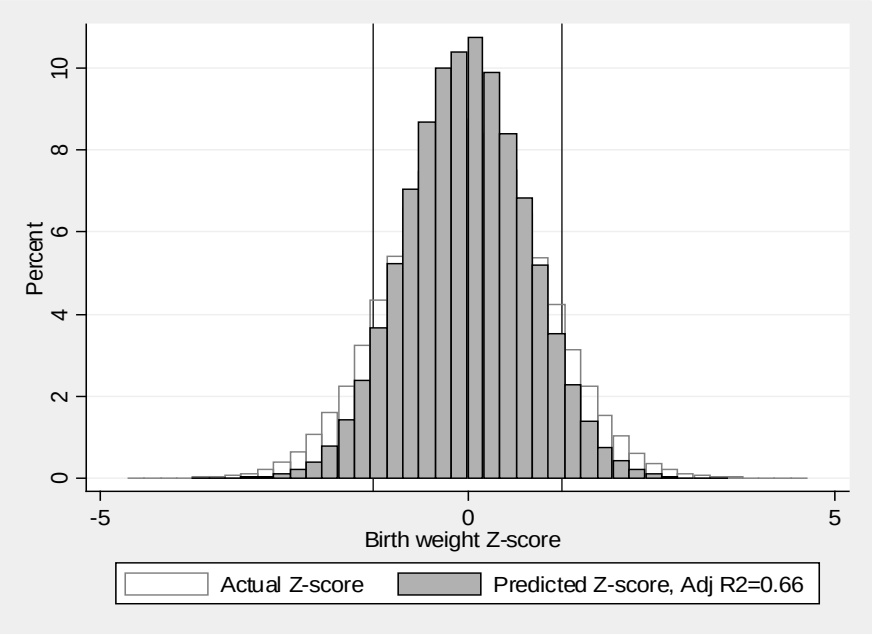


Figure 5-2. Relationship between distribution of actual birth weight Z-scores and "customized" birth weight Z-scores predicted by a customization model with a hypothetical maternal characteristics associated with birth weight Z-score by a correlation r=0.75

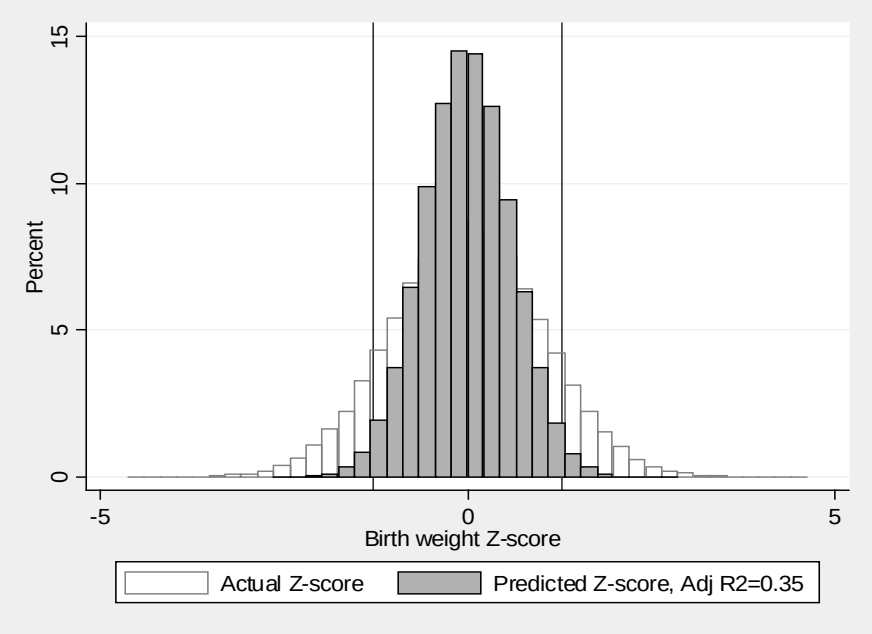


Figure 5-3. Relationship between distribution of actual birth weight Z-scores and "customized" birth weight Z-scores predicted by a customization model with a hypothetical maternal characteristics associated with birth weight Z-score by a correlation r=0.50

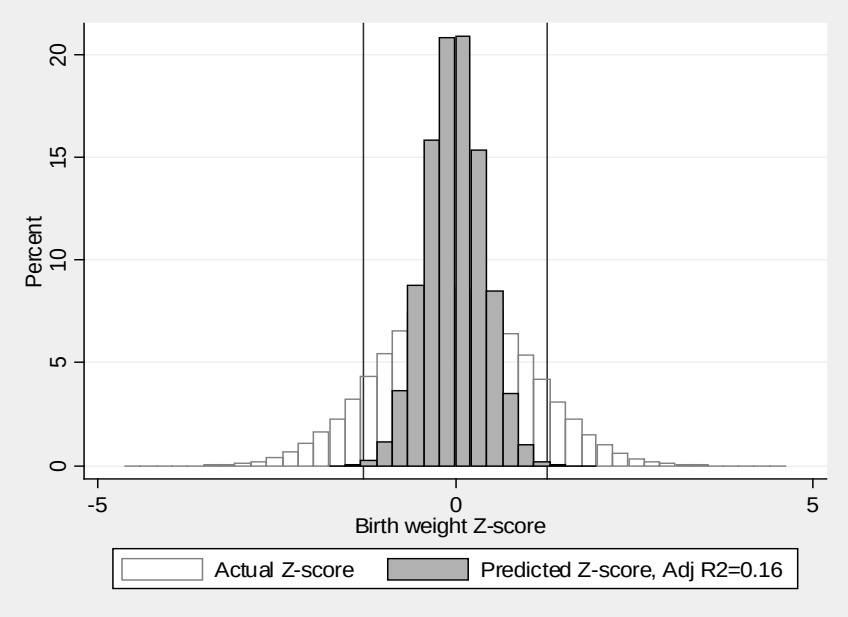


Figure 5-4. Relationship between distribution of actual birth weight Z-scores and "customized" birth weight Z-scores predicted by a customization model with a hypothetical maternal characteristics associated with birth weight Z-score by a correlation r=0.25

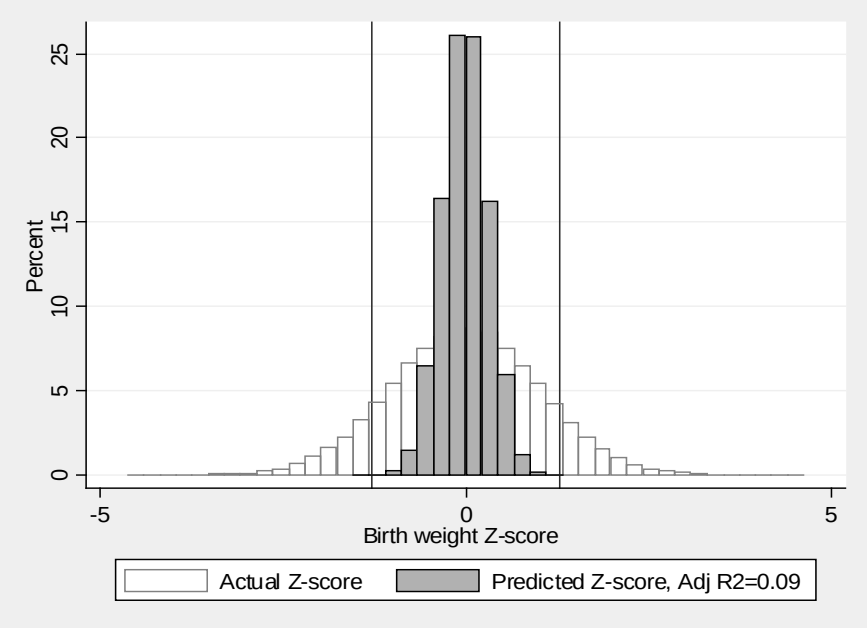


Figure 5-5. Relationship between distribution of actual birth weight Z-scores and "customized" birth weight Z-scores predicted by a customization model with only maternal characteristics used in conventional customization models included.

6. Potential alternatives: conditional fetal growth percentiles

6.1 Preamble to manuscript 3

"Conditional" fetal growth percentiles are a second alternative to conventional birthweight-for-gestational-age percentiles that have been proposed in the statistical literature. Conditional fetal growth percentiles are percentiles that are calculated given (conditional on) an infant's or fetus' weight earlier in pregnancy. Thus, they are designed to quantify the longitudinal process of fetal *growth* instead of only cross-sectional *size*. Conditional fetal growth percentiles were proposed in the statistical literature as the methodological appropriate approach to the assessment of growth,[134] and reference values for two populations (Norwegian and British) have been published from small serial ultrasound studies.[92, 135] However, whether these percentiles are actually able to improve identification of adverse perinatal outcomes related to fetal growth restriction had never been evaluated. The manuscript presented in this chapter, "The predictive ability of conditional fetal growth percentiles" (to be submitted to *Paediatric and Perinatal Epidemiology*), is the first evaluation of the clinical utility of conditional growth percentiles.

6.2 Title page & Footnotes

Title
The predictive ability of conditional fetal growth percentiles

Authors
Jennifer A HUTCHEON[1], Grace M EGELAND[2], Lucie MORIN[3], Sara J MELTZER[4,5], Geir JACOBSEN[6], Robert W PLATT[1,7]

Affiliations
McGill University, Departments of [1]Epidemiology & Biostatistics, [4]Obstetrics and Gynecology, [5]Medicine, and [6]Pediatrics

[2]McGill University, School of Dietetics and Human Nutrition

[3]University of Montreal, Department of Obstetrics and Gynecology, Sainte-Justine Hospital

[7]Department of Community Medicine and General Practice, University of Trondheim, Norway

Address for correspondence
Dr. Robert Platt
The Montreal Children's Hospital Research Institute
#205- 4060 Rue Sainte Catherine Ouest
Westmount QC
Canada H3Z 2Z3

email. robert.platt@mcgill.ca
phone. 514-934-1934 x-23288
fax. 514-412-4331

6.3 Summary

Conditional fetal growth percentiles are percentiles that are calculated taking into account (conditional on) an infant's weight earlier in pregnancy. Although they have been proposed in the statistical literature as a more methodologically appropriate method of measuring fetal growth, their ability to predict adverse perinatal outcomes due to fetal growth restriction is unknown. Using a large, unselected clinical ultrasound database at the Royal Victoria Hospital in Montreal, Canada, we calculated conditional growth percentiles for infants' weight at birth, given their weight at the time of a routine 32 or 33 week ultrasound. The risk of adverse perinatal outcome (perinatal mortality, low apgar, acidemia, or seizures/organ failure due to asphyxia) among small for gestational age infants (SGA) as established by conditional growth percentiles was calculated as well as the risk among infants classified as SGA by conventional weight-for-gestational age percentiles. Regardless of the threshold used to define SGA (5^{th}, 10^{th}, 15^{th}, 20^{th}), conditional percentiles did not appear to improve the identification of adverse perinatal outcomes compared to conventional weight-for-gestational-age charts. Further work is needed to confirm our results as well as to explore potential reasons for the lack of benefits from using a measure of growth instead of size to identify fetal growth restriction.

6.4 Introduction

A low birthweight for gestational age has long been associated with increased risk of perinatal morbidity and mortality.[4, 167] However, low weight itself is likely only a proxy for the true pathological condition of interest, poor fetal growth. It is fetuses that are small because their growth has been restricted *in utero*, not fetuses that are small but growing steadily, that are believed to be at increased risk of adverse perinatal outcomes.[15, 73] As a result, conventional approaches to identify growth-restricted infants based only on weight, such as the classification of "small-for-gestational-age" (SGA, typically defined as a birthweight below the 10th percentile)[21] will classify many small-but-healthy infants as high-risk, while failing to identify infants who did not reach their full growth potential but were not in the smallest 10 percent of the population.

A variety of approaches have been proposed to classify infants based on growth rather than size. These include the calculation of average grams gained per day,[88, 123, 124] calculation of change in weight-for-age z-scores between two time points,[120] or comparison of a current weight with that predicted through extrapolation of a weight earlier in pregnancy.[72, 168] Such approaches, however, fail to accurately account for the nature of fetal growth data, which includes both variability in trajectories *between* different fetuses (as fetuses grow at different rates) and variability *within* individual fetuses (due to biological variability from their own best-fit growth trajectory, as well as measurement error in ultrasound estimates of fetal weight).[85, 95] In order to determine if a given fetus' growth may be deviating from healthy patterns, an understanding of the amount of non-pathological variability that occurs in fetal growth measurements is crucial.

Multi-level (random effects) models have been proposed in the statistical literature as a more methodologically appropriate approach to model serial fetal weight measurements.[132, 134] In addition to providing estimates of average weights in the population at a given gestational age, these models are able to quantify the variability in growth within- and between- fetuses in the population. Most importantly, multi-level fetal growth models allow the construction of conditional fetal growth percentiles, in which the percentile assigned to a fetus' current weight is calculated taking into account (conditional on) its weight earlier in pregnancy.[134] Theoretically, conditional percentiles are the most methodologically appropriate tool to distinguish infants at

increased risk of adverse perinatal outcomes due to poor fetal growth from those with healthy growth trajectories,[95] and conditional growth percentile reference values have already been published for a United Kingdom[135] and a Norwegian population.[92] Before conditional percentiles are adopted into clinical and research use, however, an understanding of the extent to which they are actually able to identify infants at increased risk of adverse perinatal outcomes is critical. In this study, we sought to take advantage of a large, representative clinical ultrasound database to explore the predictive ability of conditional fetal growth percentiles in identifying infants at increased risk of clinical complications due to fetal growth restriction at birth, given their weight at 32-33 weeks.

6.5 Methods

Study population

The study population was drawn from women receiving antenatal care at the Royal Victoria Hospital, a McGill University teaching hospital in Montreal, Canada, between September 1996 and April 2006. Singleton pregnancies with an obstetrical ultrasound at 32-33 weeks (performed as part of routine care at our institution) were eligible for inclusion. Routine ultrasounds are available for slightly over 60% of the RVH births, which most likely reflects the preference of some clinicians at our institution to perform their ultrasounds at alternative clinics. Previous work has found that there are no significant differences in birthweight, gestational age at birth, maternal body mass index, or maternal age between births delivered at the RVH after 33 weeks who did and did not have a routine 32-33 week ultrasound record available (data available upon request).

Using maternal medical record number as a unique identifier, ultrasonographic records were linked with obstetrical and neonatal outcome data entered into the McGill Obstetric and Neonatal database (MOND), a quality-controlled clinical database that has been maintained since 1978. In the case of women who delivered more than one pregnancy at the RVH, ultrasound records were linked with the MOND based on plausibility of dates. Births between April 1, 1997 and March 31, 1998 as well as April 1, 2000 to March 31st, 2001 were not available from the MOND due to administrative reasons. Infants with congenital anomalies, and infants with missing estimates of fetal weight or outcome data were excluded. Pregnancies that were delivered at the RVH, but were referrals from outside of the hospital's source population were also excluded.

Ultrasound measurements

All ultrasound examinations were performed by certified ultrasound technicians or obstetricians with subspecialty training in ultrasonography. Fetal head circumference, abdominal circumference, and femur length were used to estimate fetal weight using the formula of Combs.[84] The formula's volumetric approach to estimation of fetal weight has been shown to have a lower systemic and total error than regression-derived EFW formulae, in particular at the extremes of the birth weight distribution.[85] Gestational age was calculated from the first day of the last normal menstrual period (LNMP). If the discrepancy between the LNMP estimate of

gestational age and an estimate obtained from early ultrasound (<20 weeks) was greater than 10 days, the latter estimate was used. If the LNMP was unknown, gestational age was based on early ultrasound estimates.

Calculation of weight percentiles

Conditional fetal growth percentiles were calculated for the infant's weight at birth given (conditional on) its estimated fetal weight at 32-33 weeks. Briefly, the conditional percentile reflects the extent to which the infant's birthweight differs from its expected birthweight had it followed a steady trajectory from its 32-33 week weight. The 50th conditional percentile would indicate that an infant had exactly the birthweight expected given its weight at 32-33 weeks, while the 10th conditional percentile would indicate that an infant was on the edge of an 80% coverage limit of its expected birthweight. The formulae used to calculate conditional percentiles[134] are provided in Appendix A. The population mean weights, estimates of the amount of non-pathological within- and between-fetus variability in weight, and of the covariance between time points required to calculate the conditional percentiles were obtained from reference values published by Owen from a UK population.[135]

Day-specific conventional weight-for-gestational-age percentiles were calculated for fetal weight at 32-33 weeks and for weight at birth using the unconditional reference values published by Owen. Although a national birthweight reference is available for the Canadian population,[12] the unconditional values published by Owen (i.e. the population sex- and age-specific means and standard deviations) were chosen to for reasons of comparability, since the use of conventional percentiles from a Canadian population and conditional percentiles from a UK population could make the former method appear better in our Canadian study population.

Composite outcome

A composite outcome of adverse perinatal events associated with fetal growth restriction was created, consisting of *any* of the following events: perinatal mortality, 5-minute apgar ≤ 3, cord pH <7.0, neonatal seizures, or organ failure (cardiac, respiratory, or renal) due to asphyxia (as diagnosed by the attending physician). All events were weighted equally.

Analysis

The risk of adverse perinatal outcome among SGA infants was calculated using conventional weight-for-gestational age percentiles (for weight at birth and estimated fetal weight at 32-33 weeks) and conditional fetal growth percentiles (for weight at birth given weight at 32 weeks). In addition to the conventional 10th percentile threshold to define SGA, SGA was also established using the 5th, 15th, and 20th percentiles at cut-off values. The relative risk of adverse perinatal outcome among SGA infants compared to non-SGA infants was calculated as:

$$\frac{\#\,adverse\ outcomes\ among\ SGA/\#\,SGA}{\#\,adverse\ outcomes\ among\ non-SGA/\#\,non-SGA}.$$

Sensitivity analyses

Several sensitivity analyses were performed. First, because identification of infants with poor fetal growth through the routine 32-33 week ultrasound could have resulted in an obstetrical intervention that prevented an adverse outcome (and thus, lower the predictive ability of 32-33 week and conditional percentiles), we included the occurrence of "delivery for fetal distress or intrauterine growth restriction (IUGR) via induction or Caesarean section" as part of our composite outcome. Delivery is the most likely treatment option for a fetus whose well-being is believed to be compromised due to growth restriction. Second, because larger infants may also be at increased risk of adverse perinatal outcomes, we restricted the reference group to infants classified as "appropriate-for-gestational-age" (AGA) by each standard (e.g. below 90th conditional percentile when SGA classified using conditional percentile standard, below 90th birthweight-for-gestational-age percentiles when using birthweight standard etc). Third, because cord pH values were missing in roughly 13% of pregnancies, and may not have been missing completely at random[143] (missing values were primarily in earlier years of the study period, and likely resulted from non-testing of infants who appeared healthy), cord pH<7 was removed as a component of the composite outcome, and risks of adverse outcome were re-calculated for the study population including infants with missing cord pH values. Finally, instead of examining the risk associated with a classification of SGA (which provides information on positive predictive value) we calculated the sensitivity and specificity of SGA as a diagnosis.

6.6 Results

A total of 11,688 singleton pregnancies with a 32-33 week ultrasound examination were available for inclusion. Removing 117 referrals from outside the hospital's source population, 784 infants with congenital anomalies, 169 infants with missing (n=168) or implausible (n=1) EFW measurements, and 1379 liveborn infants with missing cord pH values (predominantly in the years pre-2001) left n=9239 births for analysis. Descriptive characteristics of the study population are shown in Table 6-1.

In Table 6-2, the total number of adverse perinatal outcomes, as well as the number of events for each component of the composite outcome, is shown according to SGA status (<10^{th} percentile) as determined by the conditional, birthweight-for-gestational-age, and 32-33 week weight-for-gestational-age percentiles. Because some infants had more than one clinical complication of growth restriction (e.g. both a 5-minute apgar≤3 and cord pH<7), the number of events of the individual components do not sum to the total number of adverse outcomes. Although small numbers within each component do not allow definitive conclusions, constituents of the composite outcome all appeared to have reasonably similar strengths and directions of effect, and there were no obvious signs of unresponsive components.

The risks of adverse outcome among SGA infants as classified by the conditional, birthweight-for-gestational-age, and 32-33 week weight-for-gestational-age percentiles are shown in Table 6-3. Regardless of the percent threshold used to define SGA (5^{th}, 10^{th}, 15^{th}, 20^{th}), infants identified as SGA by the conventional birthweight standard were at higher risk than infants identified as SGA by either the conditional or 32-33 week weight-for-gestational-age percentiles. The relative risks with 95% confidence intervals of adverse outcome among SGA infants compared to non-SGA infants are shown in Figure 6-1a-c. Although the overlap in confidence intervals does not preclude similarities, the birthweight-for-gestational-age chart appeared to produce the highest relative risks, with point estimates above 4 at all percentile definitions of SGA. The relative risk among infants classified as SGA by the conditional percentiles were consistently lower than those of the birthweight-for-gestational-age percentiles, with estimates between 1.7 and 2.7. The point estimates for relative risks of SGA at the time of 32-33 week ultrasound also produced lower relative risks, ranging between 1.9 and 2.5. The relative risk

Table 6-1. Descriptive characteristics of the study population of 9239 singletons births at the Royal Victoria Hospital in Montreal, Canada, 1996-2006.

Maternal characteristics	**mean± SD or n(%)**
Maternal age (years)	31.8± 4.9
Maternal pre-pregnancy body mass index[1] (kg/m^2)	24.3± 5.2
Parity (% nulliparous)	5232 (49.3)
Fetal characteristics	
Sex (% male)	4641 (50.2)
Gestational age at birth (days)	276.1± 10.0
Birthweight (grams)	3423± 502
32-33 week estimated fetal weight (grams)	2080± 270
Perinatal mortality	15 (0.2)
5 minute Apgar ≤3	41 (0.4)
Cord pH <7	35 (0.4)
Neonatal seizures due to asphyxia	3 (0.03)
Organ failure due to asphyxia[2]	7 (0.1)

[1]Available for n=3583 pregnancies

[2]Organ failure defined as cardiac, respiratory, or renal failure

Table 6-2. Adverse perinatal outcomes by SGA and non-SGA status as classified by birthweight-for-gestational-age percentiles, 32-33 week estimated-fetal-weight percentiles, and conditional fetal growth percentiles among 9239 singletons births at the Royal Victoria Hospital in Montreal, Canada, 1996-2006.

	Birthweight percentiles		32-33 week estimated fetal weight percentiles		Conditional percentiles[2]	
	SGA[1] n (risk/1000)	Non-SGA n (risk/1000)	SGA[1] n (risk/1000)	Non-SGA n (risk/1000)	SGA[1] n (risk/1000)	Non-SGA n (risk/1000)
n	458	8751	754	8485	680	8559
Perinatal mortality	5 (10.9)	10 (1.1)	5 (6.6)	10 (1.2)	4 (5.9)	11 (1.3)
Cord pH<7.0	7 (15.3)	28 (3.2)	7 (9.3)	28 (3.3)	6 (8.8)	29 (3.4)
5 minute Apgar ≤3	7 (15.3)	34 (3.9)	6 (8.0)	35 (4.1)	4 (5.9)	37 (4.3)
Neonatal seizures due to asphyxia	0 (0)	3 (0.3)	1 (1.3)	2 (0.2)	0 (0)	3 (0.4)
Organ failure due to asphyxia	3 (6.6)	4 (0.5)	2 (2.7)	5 (0.6)	1 (1.5)	6 (0.7)
Composite outcome[3]	16 (34.9)	66 (7.5)	15 (20.0)	67 (7.9)	10 (14.7)	72 (8.4)

[1]Weight below the 10th percentile

[2]Weight at birth conditional on estimated fetal weight at 32 weeks

[3]Number within each component does not sum to the total number of composite outcomes since some infants had>1 adverse events.

Table 6-3. Risk of adverse perinatal outcome among SGA infants as established by birthweight-for-gestational-age percentiles, 32-33 week estimated-fetal-weight percentiles, and conditional fetal growth percentiles among 9239 singletons births at the Royal Victoria Hospital in Montreal, Canada, 1996-2006.

	<5th		<10th		<15th		<20th		>20th	
	n	Adverse outcomes n (risk/1000)	*n*	Adverse outcomes n (risk/1000)	*n*	Adverse outcomes n (risk/1000)	*n*	Adverse outcomes n (risk/1000)	*n*	Adverse outcomes n (risk/1000)
Birthweight percentile	255	10 (39.2)	458	16 (34.9)	660	21 (31.8)	909	26 (28.6)	8330	56 (6.7)
32-week weight percentile	385	8 (20.8)	754	15 (19.9)	1104	17 (15.4)	1478	22 (14.9)	7761	60 (7.7)
Conditional percentile	447	10 (22.4)	680	10 (14.7)	899	13 (14.5)	1142	18 (15.8)	8097	64 (7.9)

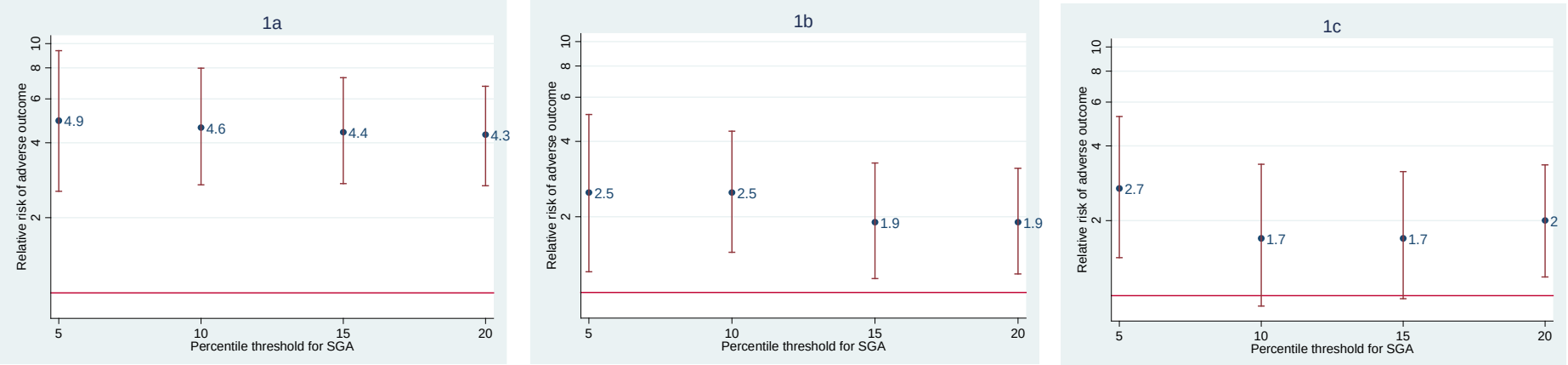


Fig 6-1a-c. Relative risk of adverse perinatal outcome, with 95% confidence intervals, as established by 1a) birth weight, 1b) 32-33 week estimated fetal weight, and 1c) conditional growth percentiles at different percentile thresholds for SGA.

among infants who were SGA by both the birthweight and the conditional percentiles (i.e. both small and poor growth) were also lower than the RRs produced by the birthweight-for-gestational age percentiles alone (RRs 4.1, 3.3, 3.5, and 4.0 for the 5^{th}, 10^{th}, 15^{th}, and 20^{th} percentile definitions of SGA, respectively).

None of the sensitivity analyses performed changed our finding that conditional percentiles were not an apparent improvement over conventional birthweight-for-gestational-age percentiles. When "delivery for fetal distress or IUGR" was included in the composite outcome, the trend in differences between the three types of percentiles was similar ($RR_{SGA(\text{birthweight percentiles})} > RR_{SGA(\text{32-week EFW percentiles})} \approx RR_{SGA(\text{conditional percentiles})}$), though confidence intervals were tighter as a result of the increased number of outcomes (n=541). Exclusion of cord pH from the composite outcome, as well as use of AGA (instead of non-SGA) as a reference group likewise did not have a major impact on our findings. Examining the sensitivity and specificity of "SGA" as classified by conditional and conventional weight-for-age standards (rather than ability to identify increased risk, which is its positive predictive value) likewise did not alter our conclusions.

6.7 Discussion

In this study, we explored the ability of conditional fetal growth percentiles to identify infants at increased risk of adverse perinatal outcomes associated with fetal growth restriction. Although conditional percentiles are theoretically a more appropriate approach to identify infants with poor *in utero* growth, our results failed to demonstrate any meaningful improvement of this approach over conventional methods.

Given the theoretical advantage of using a measure of fetal growth rather than size, our finding that conditional percentiles did not appear to improve identification of adverse perinatal outcomes was unexpected. The only other evaluation of the predictive ability of conditional percentiles that we are aware of examined the relationship between conditional z-scores for fetal abdominal area (rather than weight) and anthropometric characteristics at birth.[169] The study concluded that conditional z-scores were moderately superior to unconditional z-scores at predicting a ponderal index below the 25th percentile, but not in predicting low skinfold thickness or low mid-arm circumference/ occipito-frontal circumference ratio. Due to small sample sizes (fewer than 275 pregnancies in total), however, clinically meaningful complications of fetal growth restriction could not be examined and wide confidence intervals prevented any firm conclusions. In our study, the conditional percentiles produced lower risks and relative risks than those of conventional birthweight-for-gestational-age percentiles obtained both from our data and from earlier reports. Estimates of the relative risk of SGA (<10th percentile) have been reported to be 3.8 for stillbirth,[170] 4.5 for perinatal mortality,[114] 3.1 for a composite outcome of perinatal mortality and prolonged hospital stay,[111] and in the 2-fold range for adverse outcomes such as seizures, need for intubation, and low agpar among term infants.[40] The agreement of our study's estimates of relative risk from conventional birthweight percentiles with those in the literature lends support to our estimates of relative risk produced for conditional percentiles.

A major strength of this study was our use of a large database of ultrasounds from an unselected obstetrical population. We were therefore able to explore the predictive ability of conditional percentiles in a population 25-100 times larger than previous studies examining measures of fetal growth.[72, 120, 123, 124, 127, 169] Although our number of adverse outcomes was small, leading to wider confidence intervals, we believed that the value of having a composite outcome that was restricted to serious, rare clinical complications outweighed the gain in statistical precision that

would have been obtained had we included more subjective and less meaningful events such as neonatal intensive care unit (NICU) admissions or Caesarean sections for fetal distress.[171]

Several possible explanations for the low risks and relative risks produced by conditional percentiles should be considered. We examined the ability of conditional percentiles to identify infants at increased risk of adverse perinatal outcomes for birthweight conditioned on weight at 32-33 weeks. Although the number of grams gained per day by the fetus reaches its peak *after* 32-33 weeks of gestation,[88] it is possible that the critical period for fetal growth restriction occurs before 32 weeks. If a fetus' growth was restricted in the second trimester, but then stabilized onto a steady weight gain trajectory following the 32-33 week ultrasound, it would not be classified as "growth restricted" at birth by percentiles conditioned on its 32-33 week weight. It is also possible that conditional fetal growth percentiles may need to be modified to be able to incorporate more than 1 previous weight measurement to adequately reflect a fetus' growth trajectory in order to detect deviations from normal.[172]

We further noted that the reference values obtained from Owen's conditional percentile chart[135] appeared to underestimate the population weights at later gestational ages. Thus, while the references values fit our population reasonably well at the time of the 32-33 week ultrasound, identifying over 8% of our population as SGA (<10th percentile), by 40 weeks the 10th, 50th, and 90th percentile values predicted by Owen's reference values were all meaningfully lower than the birth weights in our population, as seen by the 10th (unconditional) percentile of Owen's standard identifying only 4.5 percent of our population as SGA. A poor fit of the reference values at later gestational ages has a major impact on the number of infants identified as SGA, since the majority of infants are born at term ages. We speculate that this discrepancy between birthweights and estimated fetal weights at later gestational ages could be the result of either poor curve fitting at the tail-ends of the ultrasound reference (as can occur with the use of higher order polynomials for gestational age rather than more flexible non-linear options such as restricted cubic splines[173, 174]) or underestimation of fetal weight at larger weights.[85] While this finding suggests that existing published conditional percentile charts likely require methodological improvements before being used in clinical practice or research, the discrepancy is nevertheless unlikely to have a major impact on the conclusions of our study. Since both the conditional percentiles and birthweight-for-gestational-age percentiles were calculated from the

same reference values, this bias would be expected to affect both types of percentiles equally. The absolute estimates for risk and relative risk produced in our study would therefore be affected, but not our study's ability to compare the risks produced by different percentiles. Furthermore, the risks of adverse outcome produced by the birthweight-for-gestational-age percentiles were consistently higher than those of the conditional percentiles even when the percent of the population classified as SGA was similar (i.e. the 15th percentile threshold for birthweight still produced a higher risk than the 10th conditional percentile threshold), suggesting that this bias in the published conditional reference values is unlikely to be the sole explanation for our findings.

Finally, estimated fetal weight is known to have 10-15% measurement error, most of which is predominantly random.[85] Given the lower relative risk also obtained with the 32-33 week weight-for-gestational age percentiles, we speculate that measurement error in estimation of fetal weight is likely a major reason for the lack of apparent benefits of conditional percentiles in identifying adverse outcomes related to fetal growth restriction.

This exploratory study failed to observe any benefits of using conditional growth percentiles instead of conventional weight-for-gestational-age percentiles in identifying adverse perinatal outcomes due to growth restriction. This could be due to error in estimated fetal weight measurements, problems in existing reference values, or a failure to adequately capture key phases of growth in the conditional percentiles. Re-assessment of this methodological approach following any future improvements in ultrasonographic imaging or ultrasound reference values is needed to help better determine the reasons for the lack of apparent benefits of conditional growth percentiles.

6.8 Appendix A. Calculation of conditional fetal growth percentiles [134]

Notation:

Y= fetal weight

Y1=fetal weight at time 1

Y2= fetal weight at time 2

X= gestational age

X1= gestational age at time 1

X2= gestational age at time 2

ε= residual error (unexplained variability)

STEP 1. Random effects model describing fetal growth

A random effects model describing fetal growth is built, with a random intercept (allowing each fetus to have its own intercept) and a random slope (allowing each fetus to have its own growth trajectory):

$$_{0}$$

where i= i^{th} fetus, j= j^{th} time point

NB For the purpose of clarity, gestational age is shown here as a linear term. In practice, a non-linear relationship between gestational age and fetal growth would be modeled (such as higher order polynomials, fractional polynomials or restricted cubic splines). [14,16,17]

STEP 2. Conditional and unconditional means and variances

The parameter estimates from the random effects model are used to calculate unconditional and conditional weights and variances for a given gestational age.

1) Calculation of unconditional (population) means and variances according to gestational age:

$$_{0}$$

$$\text{var}(\ ) \quad {}^{2}_{0} \quad {}^{2} \quad {}^{2} \quad 2 \quad {}_{0}, \quad {}^{2}$$

2) Covariance of Y1, Y2:

$$(\ 1, \ 2) \quad (\ {}_{0} \quad {}_{1} \quad 1, \ {}_{0} \quad {}_{2} \quad 2)$$
$$\quad {}^{2}_{0} \ (\ {}_{1} \quad {}_{2}) \ {}_{0}, \quad {}_{1} \ {}_{2} \quad {}^{2}$$

3) Calculation of conditional mean weight and variance at time 2 given weight at time 1:

$$(\ 2|\ 1) \quad {}_{2|1} \quad {}_{2} \quad (\ {}_{1} \quad {}_{1}) \frac{{}_{12}}{{}_{2}^{2}}$$

$$(\ 2|\ 1) \quad {}_{2}^{2} \quad \frac{{}_{12}^{2}}{{}_{2}^{2}}$$

STEP 3. Calculation of conditional percentiles

A reference interval around the expected weight at time 2, given weight at time 1 is built using the conditional variance of weight at time 2 given weight at time 1.

$$2|\ 1 \quad {}_{2|1}$$

where *z* is the standard normal deviation *z*-score

SGA infants can then be identified as those whose observed birthweight is below a desired percentile of the reference interval for expected birthweight (e.g. <10th percentile, corresponding to *z*=1.28SD).

6.9 Supplemental material for manuscript 3: Sensitivity analyses

Several sensitivity analyses were done to support the conclusion from our primary analysis that conditional fetal growth percentiles do not improve identification of infants at increased risk of fetal growth restriction compared to conventional methods. The results of these analyses, mentioned briefly in the preceding manuscript, are presented in further detail in this chapter.

a) Missing cord pH values

Of 10, 618 infants in the MOND-RVH ultrasound cohort eligible for inclusion in the study, cord pH values were missing for n=1379 (13%) of live births (stillbirths with missing cord pH values were retained in the cohort). We speculated that the cord pH values were likely not missing completely at random,[143] but instead, would be more likely to be missing from pregnancies and infants appearing to be healthy, including a birthweight classified as appropriate- for-gestational age (i.e. a cord pH would be more likely to be ordered if the infant was small, leading the clinician to investigate possible growth restriction). If infants with missing values were more likely to be appropriate-for-gestational age and from uncomplicated pregnancies, this could potentially lead to an overestimation of risk among the unexposed (non-SGA births), since infants that would have been included in the denominator would be missing. An overestimation of risk in the unexposed would in turn be expected to attenuate the relative risk of adverse outcomes among SGA infants compared to non-SGA infants.

The characteristics of women and infants with missing cord pH values were therefore compared to those with available cord pHs. The majority of the missing values were from the earlier years of the study period (pre-2000), as shown in Figure 6-2. Recall that for administrative reasons, MOND data were unavailable from the periods of Jan 1st 2007 – Jan 1st 2008 and April 1st 2000 – March 31st, 2001, explaining the gaps in the histogram during these periods. This observation suggests that missingness of values was driven by clinical practices, with clinical practices changing over the study period towards more universal testing of cord pH values.

The most notable difference between those with missing and available cord pH values was in the route of delivery of the infant. Infants with missing cord pH values were significantly less likely to be delivered via Caesarean section than those with available values (7% versus 27%,

respectively), suggesting that infants with missing values were more likely to come from uncomplicated, low-risk pregnancies. Mothers of infants with missing cord pH values were younger (by an average of 7 months, p<0.0001) and had lower pre-pregnancy weights (1.5kg lighter, p=0.005) than mothers of infants for whom cord pH was available. There was no significant difference in birthweight-for-age according to missing cord pH status (birthweight ratio difference of 0.002, p=0.57), but infants with missing cord values were slightly younger) than those with available values (2 days, p<0.0001, and were correspondingly lighter (51 grams, p=0.0003).

Because of the potential for bias from cord pH values that were not missing completely at random, we repeated our analyses excluding "cord pH <7" as a component of the composite outcome, and re-calculated risks and relative risks of the new composite outcome for the entire 10,618 eligible infants (i.e. including those infants with missing cord pH values). Exclusion of "cord pH <7" as a component of the composite outcome resulted in fewer total adverse outcomes (n=54), and as a result, wider confidence intervals. The risks and relative risks, however revealed the same trends as seen in our primary analysis, with a classification of SGA by the conventional birthweight percentiles producing the highest risks and relative risks of adverse perinatal outcomes (Table 6-4, Figure 6-3a-c).

b) Inclusion of "Delivery for fetal distress or intrauterine growth restriction" in composite outcome

Because of the observational nature of this study's data, it is important to consider that perinatal outcomes may have been altered as a result of the 32-33 week ultrasound. If the results of the 32-33 week ultrasound revealed an infant to be at increased risk due to growth restriction, for example, the pregnancy would subsequently likely be more closely monitored, and the fetus delivered if signs of fetal demise became apparent. Thus, an infant may have had a slowing growth trajectory (and as a result, a low conditional fetal growth percentile), but no adverse perinatal outcome due to a timely obstetrical intervention. Any such obstetrical intervention would result in a decreased risk of adverse events among infants classified as SGA. Since the primary treatment option for intrauterine growth restriction is delivery, we therefore included "Caesarean section/induction for fetal distress or intrauterine growth restriction" as a component of our composite outcome. We included only pregnancies where "fetal distress" or

"intrauterine growth restriction" was either the primary indication (MOND1 period) or the sole indication (MOND2 period). Although this approach to categorization of reasons for delivery is imperfect, we felt that it was a most conservative approach to identify fetuses that were truly threatened as a result of growth restriction, rather than including all medically indicated deliveries where "fetal distress" or "intrauterine growth restriction" was listed, but as one of up to 5 indications.

The inclusion of "delivery for fetal distress or intrauterine growth restriction" as a component of the composite outcome resulted in a large increase in the number of adverse outcomes (n=541), and as a result, correspondingly more narrow 95% confidence intervals. As in our primary analysis, the risks and relative risks obtained from the conventional birthweight-for-gestational-age percentiles were higher across all thresholds for SGA than the 32-33 week ultrasound percentiles or conditional percentiles (Table 6-5, Figure 6-4a-c). Unlike the results seen with our primary composite outcome, the inclusion of "Delivery for fetal distress or IUGR" in the composite outcome resulted in a trend of increasing relative risk with decreasing SGA threshold percentile for the 32-33 week ultrasound percentiles. This likely reflects the influence of estimated fetal weight on clinical decision making, with infants identified as small by the ultrasound examination being more likely to be delivered because of concerns over potential fetal demise.

c) Restriction of referent group to "appropriate-for-gestational-age" infants

The relationship between birthweight-for-gestational-age and adverse outcomes has been observed to follow an inverted 'J'-shape, with those at the extremes of the birthweight continuum being at increased risk. Although the risk among small infants is markedly higher than the risk among large infants, it is possible than failure to exclude larger-than-average infants from the referent group of "non-SGA" infants could result in an increased risk among the unexposed group, and as a result, attenuate the apparent relative risk of adverse outcomes among SGA infants. We therefore repeated our calculations of risk and relative risk excluding infants that were large-for-gestational-age (LGA, defined as >90th percentile according to each standard) from the referent group. The results, shown in Table 6-6 and Figure 6-5a-c, demonstrate that using "appropriate-for-gestational-age" as a reference group instead of "non-SGA" had little impact on our conclusions. This result helps validate our choice of a composite

outcome, which was intended to target infants with adverse outcomes resulting from fetal growth restriction and should not occur in infants with fetal overgrowth in higher than average levels.

d) Sensitivity and specificity

Finally, although conventionally, the classification of SGA has been considered a "risk factor" for adverse perinatal outcomes and the goal of research has been to quantify the degree of increased risk associated with SGA infants, birthweight-for-gestational-age percentiles can alternatively be viewed as a screening tool, with "SGA" corresponding to a positive test result. We therefore examined the sensitivity and specificity of "SGA" as a diagnostic tool to classify adverse perinatal outcomes due to fetal growth restriction (Table6-7). Receiver operating characteristic (ROC) curves were then built according to standard methods, and the area under the curve calculated (Figure 6-6a-c).[175] While the conventional birthweight charts consistently had the highest sensitivity and specificity across each of the percentile thresholds assessed, the low area under the curve of all 3 approaches highlights that while SGA infants may be at increased risk of adverse perinatal outcomes, "small-for-gestational-age" is not clinically useful diagnosis for adverse perinatal outcomes.

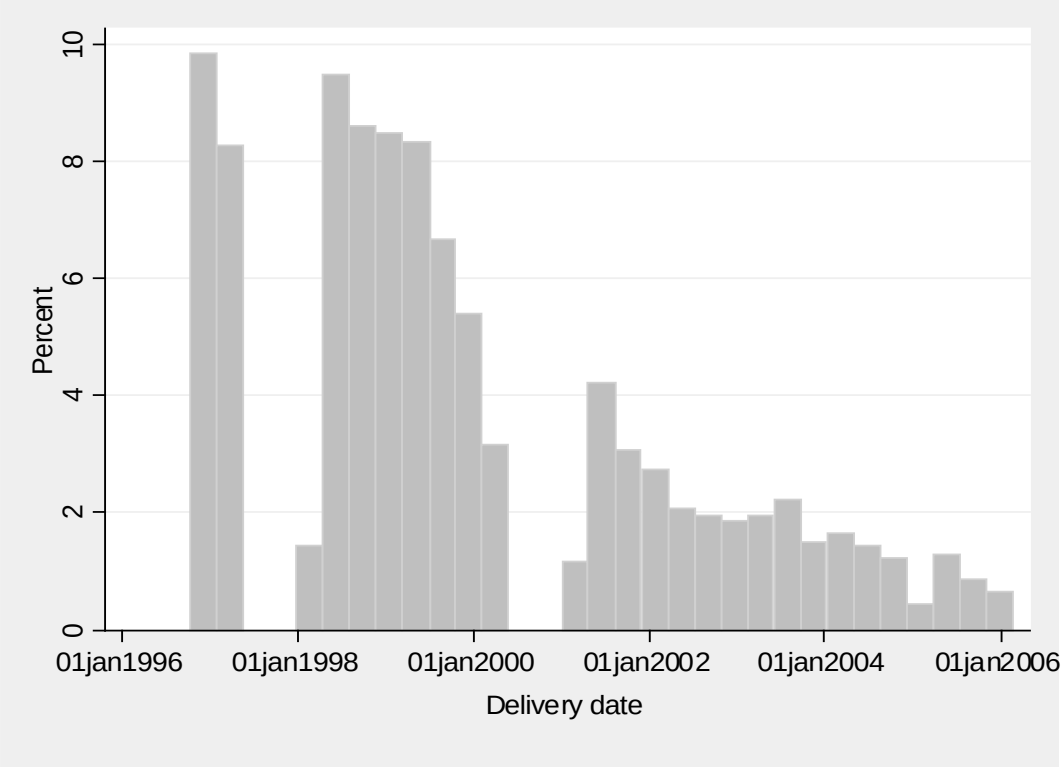


Figure 6-2. Distribution of delivery dates of infants with missing cord pH values

Table 6-4. Risk of adverse perinatal outcome among SGA infants as established by conditional fetal growth percentiles, birthweight-for-gestational-age percentiles, and 32-33 week estimated-fetal-weight-for-gestational-age percentiles, with cord pH excluded from composite outcome.

	<5th		<10th		<15th		<20th		>20th	
	n	Adverse outcomes n (risk/1000)	*n*	Adverse outcomes n (risk/1000)	*N*	Adverse outcomes n (risk/1000)	*N*	Adverse outcomes n (risk/1000)	*n*	Adverse outcomes n (risk/1000)
Birthweight percentile	291	7 (24.1)	513	10 (19.5)	750	14 (18.7)	1020	18 (17.6)	9598	36 (3.8)
32-week weight percentile	448	5 (11.2)	866	10 (11.5)	1271	11 (8.7)	1698	15 (8.8)	8920	39 (4.4)
Conditional percentile	515	5 (9.7)	780	5 (6.4)	1036	8 (7.7)	1315	12 (9.1)	9303	42 (4.5)

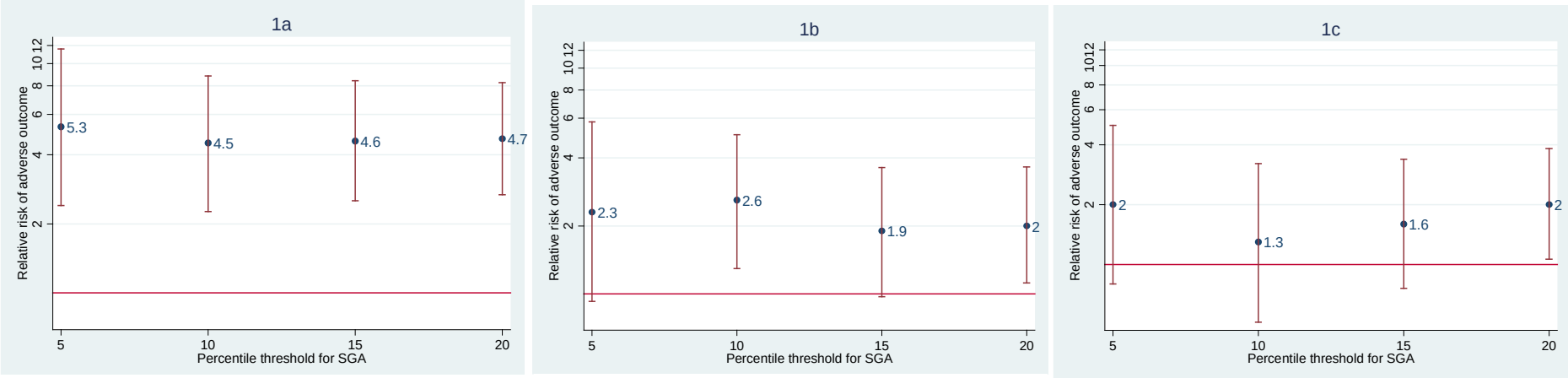


Figure 6-3a-c. Relative risk of adverse perinatal outcome as established by 1a) birth weight, 1b) 32-33 week estimated fetal weight, and 1c) conditional growth percentiles at different percentile thresholds for SGA with cord pH excluded as a component of the composite outcome.

Inclusion of "Delivery for fetal distress or intrauterine growth restriction" in composite outcome

Table 6-5. Risk of adverse perinatal outcome among SGA infants as established by conditional fetal growth percentiles, birthweight-for-gestational-age percentiles, and 32-33 week estimated-fetal-weight-for-gestational-age percentiles, with "Delivery for fetal distress or intrauterine growth restriction" included in composite outcome.

	<5th		<10th		<15th		<20th		>20th	
	n	Adverse outcomes n (risk/1000)	*n*	Adverse outcomes n (risk/1000)	*n*	Adverse outcomes n (risk/1000)	*n*	Adverse outcomes n (risk/1000)	*n*	Adverse outcomes n (risk/1000)
Birthweight percentile	255	103 (404)	458	154 (336)	660	196 (297)	909	239 (263)	8330	302 (36)
32-week weight percentile	385	120 (312)	754	190 (252)	1104	233 (211)	1478	262 (177)	7761	279 (36)
Conditional percentile	447	75 (168)	680	99 (156)	899	123 (137)	1142	146 (128)	8097	395 (45)

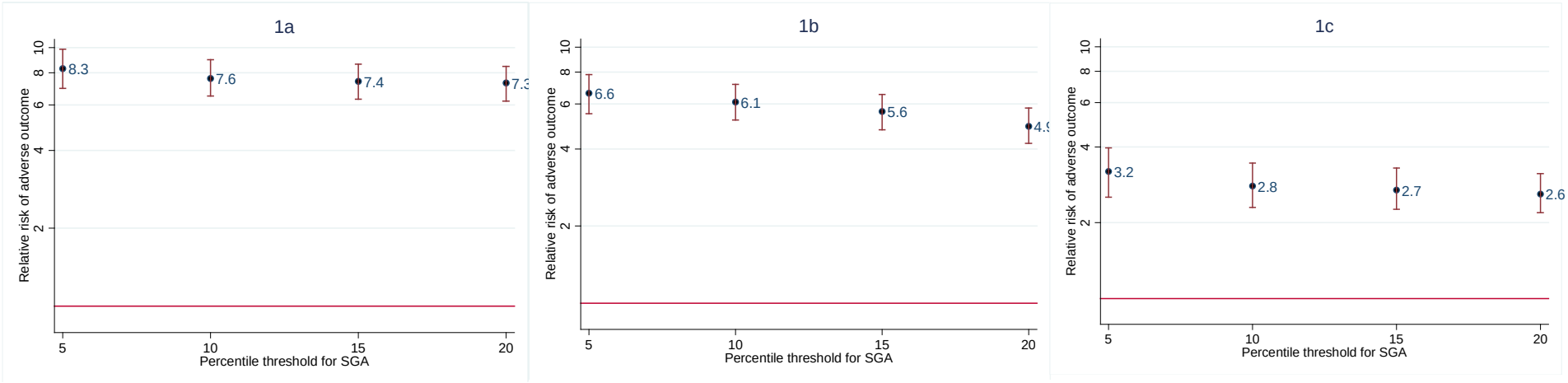


Figure 6-4a-c. Relative risk of adverse perinatal outcome as established by 1a) birth weight, 1b) 32-33 week estimated fetal weight, and 1c) conditional growth percentiles at different percentile thresholds for SGA with "delivery for fetal distress or intrauterine growth restriction" included in composite outcome.

AGA as referent group

Table 6-6. Risk of adverse perinatal outcome among SGA infants as established by conditional fetal growth percentiles, birthweight-for-gestational-age percentiles, and 32-33 week estimated-fetal-weight-for-gestational-age percentiles, with large-for-gestational-age (>90th percentile) infants excluded from referent group.

	<5th		<10th		<15th		<20th		20th to 90th	
	n	Adverse outcomes n (risk/1000)	*n*	Adverse outcomes n (risk/1000)	*n*	Adverse outcomes n (risk/1000)	*n*	Adverse outcomes n (risk/1000)	*n*	Adverse outcomes n (risk/1000)
Birthweight percentile	255	10 (39.2)	458	16 (34.9)	660	21 (31.8)	909	26 (28.6)	6218	41 (6.6)
32-week weight percentile	385	8 (20.8)	754	15 (19.9)	1104	17 (15.4)	1478	22 (14.9)	6526	51 (7.8)
Conditional percentile	447	10 (22.4)	680	10 (14.7)	899	13 (14.5)	1142	18 (15.8)	5231	48 (9.2)

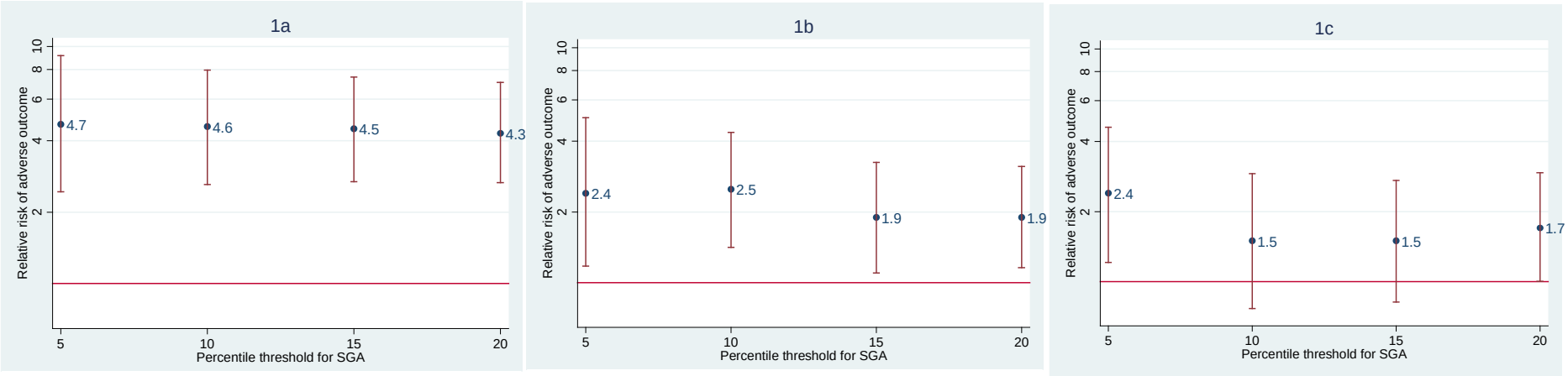


Figure 6-5a-c. Relative risk of adverse perinatal outcome as established by 1a) birth weight, 1b) 32-33 week estimated fetal weight, and 1c) conditional growth percentiles at different percentile thresholds for SGA with large-for-gestational age (>90th percentile) excluded from reference group.

Sensitivity and specificity

Table 6-7. Sensitivity, specificity, and area under the receiver-operating characteristic curve of birthweight, 32-week estimated fetal weight, and conditional fetal growth percentiles.

		<5th percentile	<10th percentile	<15th percentile	<20th percentile	AUC*
Birthweight percentiles	Sensitivity (%)	31.7	25.6	19.5	12.2	0.61
	Specificity (%)	90.4	93.0	95.2	97.3	
32-week estimated fetal weight percentiles	Sensitivity (%)	26.8	20.7	18.3	9.8	0.56
	Specificity (%)	84.1	88.1	91.9	95.9	
Conditional percentiles	Sensitivity (%)	22.0	15.9	12.2	12.2	0.55
	Specificity (%)	87.7	90.3	92.7	95.2	

* AUC= area under the receiver operating characteristic (ROC) curve

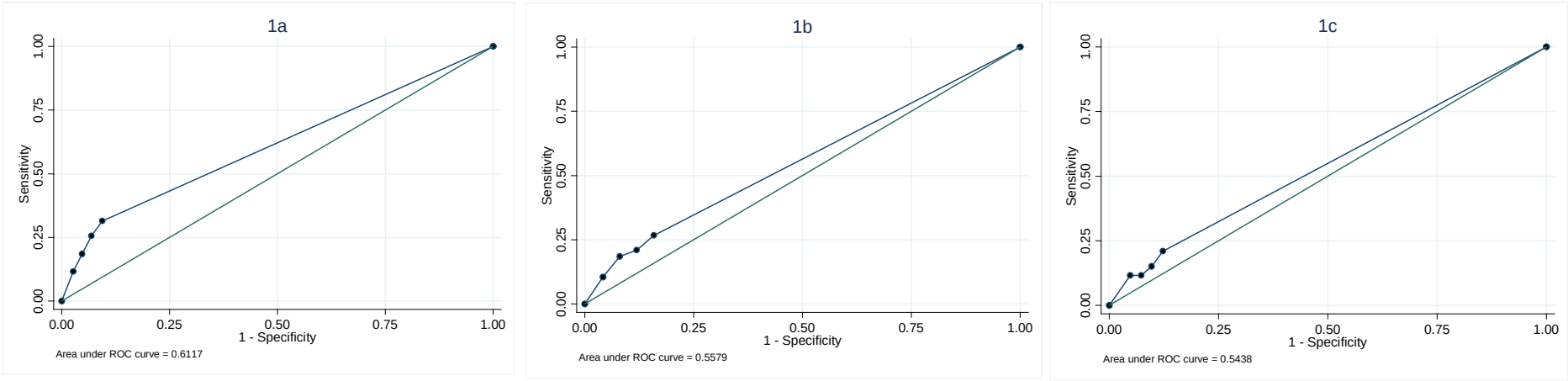


Figure 6-6a-c. Receiver-operating characteristic (ROC) curves for 1a) birth weight, 1b) 32-33 week estimated fetal weight, and 1c) conditional growth percentiles in the identification of adverse perinatal outcomes due to fetal growth restriction.

7. Conclusions

The objectives of this thesis were 1) to evaluate the potential for bias arising from the use of conventional birthweight percentiles in the study of fetal growth restriction, and 2) to evaluate alternatives to the conventional definition of "small-for-gestational age" that have been proposed in the literature, namely customized birthweight percentiles and conditional fetal growth percentiles. In the first manuscript of the thesis, we illustrated how conventional birthweight charts have serious shortcomings at preterm ages due to the missing data from fetuses that remain *in utero.* We demonstrated that these missing data introduce a non-trivial degree of bias to epidemiologic studies of fetal growth restriction that use "small-for-gestational-age" as an outcome definition, and concluded with the recommendation that classifications of "small-for-gestational- age" should be based on the 10^{th} percentile of all fetuses at risk of being small at a given gestational age, not the 10^{th} percentile of birthweights.

This recommendation, however, was based solely on theoretical grounds from basic epidemiologic first principles. Empirical evidence to support our recommendation can be found in the results of the second thesis manuscript. Although the manuscript's primary goal was to evaluate the benefits of customizing for maternal characteristics, the comparison of the customized standard to both intrauterine and birthweight standards also provided an evaluation of the latter two standards' relative ability to identify infants at increased risk of perinatal mortality. We found that at preterm ages, the 10^{th} percentile of the intrauterine population was better able to identify risk for both stillbirth and early neonatal mortality than the 10^{th} percentile of the weights of livebirths. Since at early preterm ages, the intrauterine population is virtually identical to the total at-risk population (because the number of infants in the cohort born at early gestational ages is negligible), these results illustrate the need to define "small-for-gestational-age" as the 10^{th} percentile of the total cohort, not the 10^{th} percentile of birthweights. As term approached, the relative risks produced by the intrauterine and birthweight standards became more similar, and within term ages, the relative risks of perinatal mortality produced by the classification of SGA by the intrauterine standard were not meaningfully different than those produced by the conventional birthweight chart. This result also supports our recommendations, because at term ages, there is no (or minimal) missing data bias in

conventional birthweight charts, so no improvement in identification of risk for perinatal mortality from the use of a standard derived from intrauterine weights would be expected. Our comparison of intrauterine, birthweight, and customized percentiles thus provided evidence of the inappropriateness of using conventional birthweight charts to define SGA, and of the benefits of basing SGA thresholds on the weight distribution of the total cohort at-risk.

The evaluation of alternatives to conventional weight-for-gestational-age charts (customized percentiles and conditional fetal growth percentiles) failed to demonstrate any advantages over conventional methods. The second thesis manuscript showed that customizing birth weight percentiles by adjusting for maternal characteristics provided no meaningful advantages over a population weight-for-age approach once the bias at preterm ages in conventional birthweight charts had been corrected through the use of an intrauterine standard. Conditional fetal growth percentiles also did not appear to be an improvement over conventional approaches (manuscript 3), although this methodology deserves re-evaluation once improved conditional reference values and ultrasonographic imaging are available.

While the alternative approaches to the classification of fetal growth assessed in this thesis may not have been found to be improvements over conventional weight-for-gestational-age percentiles, the thesis has highlighted the need to resolve the bias in conventional percentiles from missing intrauterine weights at preterm ages. A *perinatal* weight reference that is representative of the weights of all fetuses at risk of being small at a given gestational age, regardless of their birth status, is needed. Although serial ultrasound weight measurements from a large, population-based sample are not available at present to create this perinatal weight reference, data on all stages of fetal growth do exist, albeit in fragmented sources and each with its own set of limitations. Research studies with repeated serial ultrasounds provide an estimate of the "shape" of fetal growth trajectories throughout gestation, but are based on small sample sizes (generally several hundred) and potentially non-representative research study populations.[16, 91, 92, 176, 177] Clinical ultrasound databases such as that of the Royal Victoria Hospital provide much larger sample sizes, and are representative of a general obstetrical population, but only provide weights at a single time window of the third trimester (the 32-33 week period). Ultrasonographic estimates of fetal weight are also prone to measurement error.[85] Finally, population birth registries provide large, representative samples for term births,

but are missing the weights of ongoing pregnancies at preterm ages. To produce an unbiased perinatal weight reference, the information in these different datasources would need to be combined, triangulating available information to overcome the limitations of each type of data.

Work is currently underway in our research group to create a simulated pregnancy cohort for the study of fetal growth and infant outcomes. This simulated cohort will be created by combining data from the multiple datasources described above, including serial ultrasound research studies, clinical ultrasound databases, and population birth registries. Estimates of the "shape" of fetal growth curves can be obtained from serial ultrasound research studies, but the absolute values for mean weights and variances will need to be adjusted to ensure that they are representative of general obstetrical populations, using clinical ultrasound databases of routine ultrasounds at preterm ages, and population birth registries at later gestational ages. Further, information on the strength and direction of measurement error in estimated fetal weight values obtained from EFW formulae validation studies will be used to correct the means and variances of population weights obtained from ultrasound data. By producing a model for fetal growth that reflects the weights of the entire cohort at risk at each gestational age, a new perinatal weight reference will be produced. Future work to validate this reference and examine its ability to predict adverse perinatal outcomes due to growth restriction is needed, but this new simulated perinatal weight reference has the potential to be an important contribution to perinatal epidemiology by improving the clinical prediction of high-risk infants as well as creating an unbiased case definition of "small-for-gestational-age" for research purposes.

8. References

1. Zhang X, Platt RW, Cnattingius S, Joseph KS, Kramer MS. The use of customised versus population-based birthweight standards in predicting perinatal mortality. BJOG 2007;114:474-7.
2. Schlesinger ER, Allaway NC. The combined effect of birth weight and length of gestation on neonatal mortality among single premature births. Pediatrics 1955;15:698-704.
3. Rosenberg A. The IUGR Newborn. Semin Perinatol 2008;32:219-24.
4. Wilcox AJ, Skjaerven R. Birth weight and perinatal mortality: the effect of gestational age. Am J Public Health 1992;82:378-82.
5. World Health Organization. Expert group on prematurity. Technical report, series 27. Geneva: World Health Organization; 1950.
6. W.E.N. Categorization by weight and gestational age of the infant at birth. The Journal of pediatrics 1967;71:309-10.
7. Battaglia FC, Lubchenco LO. A practical classification of newborn infants by weight and gestational age. J Pediatr 1967;71:159-63.
8. Lubchenco LO, Hansman C, Dressler M, Boyd E. Intrauterine Growth as Estimated from Liveborn Birth-Weight Data at 24 to 42 Weeks of Gestation. Pediatrics 1963;32:793-800.
9. Alexander GR, Himes JH, Kaufman RB, Mor J, Kogan M. A United States national reference for fetal growth. Obstet Gynecol 1996;87:163-8.
10. Bonellie S, Chalmers J, Gray R, Greer I, Jarvis S, Williams C. Centile charts for birthweight for gestational age for Scottish singleton births. BMC Pregnancy Childbirth 2008;8:5.
11. Kallen B. A birth weight for gestational age standard based on data in the Swedish Medical Birth Registry, 1985-1989. Eur J Epidemiol 1995;11:601-6.
12. Kramer MS, Platt RW, Wen SW, et al. A new and improved population-based Canadian reference for birth weight for gestational age. Pediatrics 2001;108:E35.
13. Skjaerven R, Gjessing HK, Bakketeig LS. Birthweight by gestational age in Norway. Acta Obstet Gynecol Scand 2000;79:440-9.
14. Zhang J, Bowes WA, Jr. Birth-weight-for-gestational-age patterns by race, sex, and parity in the United States population. Obstet Gynecol 1995;86:200-8.
15. Altman DG, Hytten FE. Intrauterine growth retardation: let's be clear about it. Br J Obstet Gynaecol 1989;96:1127-32.
16. Marsal K, Persson PH, Larsen T, Lilja H, Selbing A, Sultan B. Intrauterine growth curves based on ultrasonically estimated foetal weights. Acta Paediatr 1996;85:843-8.
17. Ott WJ. Intrauterine growth retardation and preterm delivery. Am J Obstet Gynecol 1993;168:1710-5; discussion 5-7.
18. Secher NJ, Kern Hansen P, Thomsen BL, Keiding N. Growth retardation in preterm infants. Br J Obstet Gynaecol 1987;94:115-20.
19. Weiner CP, Sabbagha RE, Vaisrub N, Depp R. A hypothetical model suggesting suboptimal intrauterine growth in infants delivered preterm. Obstet Gynecol 1985;65:323-6.
20. Khoury MJ, Erickson JD, Cordero JF, McCarthy BJ. Congenital malformations and intrauterine growth retardation: a population study. Pediatrics 1988;82:83-90.
21. Resnik R. Intrauterine growth restriction. Obstet Gynecol 2002;99:490-6.

22. Gabbe SG, Niebyl JR, Simpson JL. Obstetrics : normal and problem pregnancies. 4th ed. New York: Churchill Livingstone; 2002.
23. Hendrix N, Berghella V. Non-placental causes of intrauterine growth restriction. Semin Perinatol 2008;32:161-5.
24. Maulik D. Fetal growth restriction: the etiology. Clin Obstet Gynecol 2006;49:228-35.
25. Houlton MC, Marivate M, Philpott RH. The prediction of fetal growth retardation in twin pregnancy. Br J Obstet Gynaecol 1981;88:264-73.
26. Heinonen S, Taipale P, Saarikoski S. Weights of placentae from small-for-gestational age infants revisited. Placenta 2001;22:399-404.
27. Maulik D, Frances Evans J, Ragolia L. Fetal growth restriction: pathogenic mechanisms. Clin Obstet Gynecol 2006;49:219-27.
28. Vik T, Vatten L, Jacobsen G, Bakketeig LS. Prenatal growth in symmetric and asymmetric small-for-gestational-age infants. Early Hum Dev 1997;48:167-76.
29. Kramer MS, Olivier M, McLean FH, Willis DM, Usher RH. Impact of intrauterine growth retardation and body proportionality on fetal and neonatal outcome. Pediatrics 1990;86:707-13.
30. Bauer MK, Harding JE, Bassett NS, et al. Fetal growth and placental function. Mol Cell Endocrinol 1998;140:115-20.
31. Groff JL, Gropper SAS. Advanced nutrition and human metabolism. 3rd ed. Belmont, CA: West/Wadsworth; 2000.
32. Baschat AA. Fetal responses to placental insufficiency: an update. BJOG 2004;111:1031-41.
33. Hooper SB. Fetal metabolic responses to hypoxia. Reprod Fertil Dev 1995;7:527-38.
34. Pallotto EK, Kilbride HW. Perinatal outcome and later implications of intrauterine growth restriction. Clin Obstet Gynecol 2006;49:257-69.
35. McKeown T, Gibson JR. Observations of all births (23,970) in Birmingham, 1947. IV. Premature birth. Br Med J 1951;2:513-7.
36. Wegman ME. Weight at birth and survival of the newborn. Pediatrics 1954;14:396-400.
37. Arnold CC, Kramer MS, Hobbs CA, McLean FH, Usher RH. Very low birth weight: a problematic cohort for epidemiologic studies of very small or immature neonates. Am J Epidemiol 1991;134:604-13.
38. Yerushalmy J. The classification of newborn infants by birth weight and gestational age. The Journal of pediatrics 1967;71:164-72.
39. Janssen P, A. , Thiessen P, Klein MC, Whitfield M, MacNab YC, Cullis-Kuhl S. Standards for the measurement of birth weight, length and head circumference at term in neonates of European, Chinese and South Asian ancestry. Open Medicine 2007.
40. McIntire DD, Bloom SL, Casey BM, Leveno KJ. Birth weight in relation to morbidity and mortality among newborn infants. N Engl J Med 1999;340:1234-8.
41. Seeds JW, Peng T. Impaired growth and risk of fetal death: is the tenth percentile the appropriate standard? Am J Obstet Gynecol 1998;178:658-69.
42. Catov JM, Nohr EA, Olsen J, Ness RB. Chronic hypertension related to risk for preterm and term small for gestational age births. Obstet Gynecol 2008;112:290-6.
43. Boulet SL, Alexander GR, Salihu HM, Kirby RS, Carlo WA. Fetal growth risk curves: defining levels of fetal growth restriction by neonatal death risk. Am J Obstet Gynecol 2006;195:1571-7.
44. Joseph KS, Fahey J, Platt RW, et al. An outcome-based approach for the creation of fetal growth standards: do singletons and twins need separate standards? Am J Epidemiol 2009;in press.

45. Infante-Rivard C, Rivard GE, Yotov WV, et al. Absence of association of thrombophilia polymorphisms with intrauterine growth restriction. N Engl J Med 2002;347:19-25.
46. Ahluwalia IB, Grummer-Strawn L, Scanlon KS. Exposure to environmental tobacco smoke and birth outcome: increased effects on pregnant women aged 30 years or older. Am J Epidemiol 1997;146:42-7.
47. Buchbinder A, Sibai BM, Caritis S, et al. Adverse perinatal outcomes are significantly higher in severe gestational hypertension than in mild preeclampsia. Am J Obstet Gynecol 2002;186:66-71.
48. Liu S, Wen SW, Demissie K, Marcoux S, Kramer MS. Maternal asthma and pregnancy outcomes: a retrospective cohort study. Am J Obstet Gynecol 2001;184:90-6.
49. Arends NJ, Boonstra VH, Duivenvoorden HJ, Hofman PL, Cutfield WS, Hokken-Koelega AC. Reduced insulin sensitivity and the presence of cardiovascular risk factors in short prepubertal children born small for gestational age (SGA). Clin Endocrinol (Oxf) 2005;62:44-50.
50. Hofman PL, Cutfield WS, Robinson EM, et al. Insulin resistance in short children with intrauterine growth retardation. J Clin Endocrinol Metab 1997;82:402-6.
51. Robinson S, Walton RJ, Clark PM, Barker DJ, Hales CN, Osmond C. The relation of fetal growth to plasma glucose in young men. Diabetologia 1992;35:444-6.
52. Willemsen RH, de Kort SW, van der Kaay DC, Hokken-Koelega AC. Independent effects of prematurity on metabolic and cardiovascular risk factors in short small-for-gestational-age children. J Clin Endocrinol Metab 2008;93:452-8.
53. Birth weight for gestational age. 2004. (Accessed September 21, 2008, 2008, at http://www.phac-aspc.gc.ca/rhs-ssg/bwga-pnag/index-eng.php.)
54. Kramer MS, McLean FH, Boyd ME, Usher RH. The validity of gestational age estimation by menstrual dating in term, preterm, and postterm gestations. JAMA 1988;260:3306-8.
55. David RJ. The quality and completeness of birthweight and gestational age data in computerized birth files. Am J Public Health 1980;70:964-73.
56. Selbing A, Fjallbrant B. Accuracy of conceptual age estimation from fetal crown-rump length. J Clin Ultrasound 1984;12:343-6.
57. Hadlock FP, Harrist RB, Martinez-Poyer J. How accurate is second trimester fetal dating? J Ultrasound Med 1991;10:557-61.
58. Blondel B, Morin I, Platt RW, Kramer MS, Usher R, Breart G. Algorithms for combining menstrual and ultrasound estimates of gestational age: consequences for rates of preterm and postterm birth. BJOG 2002;109:718-20.
59. Joseph KS, Huang L, Liu S, et al. Reconciling the high rates of preterm and postterm birth in the United States. Obstet Gynecol 2007;109:813-22.
60. Morin I, Morin L, Zhang X, et al. Determinants and consequences of discrepancies in menstrual and ultrasonographic gestational age estimates. BJOG 2005;112:145-52.
61. Platt RW, Abrahamowicz M, Kramer MS, et al. Detecting and eliminating erroneous gestational ages: a normal mixture model. Stat Med 2001;20:3491-503.
62. Burkhardt T, Schaffer L, Zimmermann R, Kurmanavicius J. Newborn weight charts underestimate the incidence of low birthweight in preterm infants. Am J Obstet Gynecol 2008;199:139 e1-6.
63. Zeitlin J, Ancel PY, Saurel-Cubizolles MJ, Papiernik E. The relationship between intrauterine growth restriction and preterm delivery: an empirical approach using data from a European case-control study. BJOG 2000;107:750-8.
64. Morken NH, Kallen K, Jacobsson B. Fetal growth and onset of delivery: a nationwide population-based study of preterm infants. Am J Obstet Gynecol 2006;195:154-61.

65. Hediger ML, Scholl TO, Schall JI, Miller LW, Fischer RL. Fetal growth and the etiology of preterm delivery. Obstet Gynecol 1995;85:175-82.
66. Zhang J, Sundaram R, Sun W, Troendle J. Fetal Growth and Timing of Parturition in Humans. Am J Epidemiol 2008.
67. Ewigman BG, Crane JP, Frigoletto FD, LeFevre ML, Bain RP, McNellis D. Effect of prenatal ultrasound screening on perinatal outcome. RADIUS Study Group. N Engl J Med 1993;329:821-7.
68. Papiernik E. Fetal growth retardation: a limit for the further reduction of preterm births. Matern Child Health J 1999;3:63-9.
69. Joseph KS, Demissie K, Kramer MS. Obstetric intervention, stillbirth, and preterm birth. Semin Perinatol 2002;26:250-9.
70. Wilcox AJ. Birth weight, gestation, and the fetal growth curve. Am J Obstet Gynecol 1981;139:863-7.
71. Soothill PW, Bobrow CS, Holmes R. Small for gestational age is not a diagnosis. Ultrasound Obstet Gynecol 1999;13:225-8.
72. Danielian PJ, Allman AC, Steer PJ. Is obstetric and neonatal outcome worse in fetuses who fail to reach their own growth potential? Br J Obstet Gynaecol 1992;99:452-4.
73. Patterson RM, Pouliot MR. Neonatal morphometrics and perinatal outcome: who is growth retarded? Am J Obstet Gynecol 1987;157:691-3.
74. McNay MB, Fleming JE. Forty years of obstetric ultrasound 1957-1997: from A-scope to three dimensions. Ultrasound Med Biol 1999;25:3-56.
75. Campbell S, Wilkin D. Ultrasonic measurement of fetal abdomen circumference in the estimation of fetal weight. Br J Obstet Gynaecol 1975;82:689-97.
76. Shepard MJ, Richards VA, Berkowitz RL, Warsof SL, Hobbins JC. An evaluation of two equations for predicting fetal weight by ultrasound. Am J Obstet Gynecol 1982;142:47-54.
77. Warsof SL, Gohari P, Berkowitz RL, Hobbins JC. The estimation of fetal weight by computer-assisted analysis. Am J Obstet Gynecol 1977;128:881-92.
78. Eik-Nes SH, Grottum P. Estimation of fetal weight by ultrasound measurement. I. Development of a new formula. Acta Obstet Gynecol Scand 1982;61:299-305.
79. Hadlock FP, Harrist RB, Sharman RS, Deter RL, Park SK. Estimation of fetal weight with the use of head, body, and femur measurements--a prospective study. Am J Obstet Gynecol 1985;151:333-7.
80. Persson PH, Weldner BM. Intra-uterine weight curves obtained by ultrasound. Acta Obstet Gynecol Scand 1986;65:169-73.
81. Rose BI, McCallum WD. A simplified method for estimating fetal weight using ultrasound measurements. Obstet Gynecol 1987;69:671-5.
82. Sabbagha RE, Minogue J, Tamura RK, Hungerford SA. Estimation of birth weight by use of ultrasonographic formulas targeted to large-, appropriate-, and small-for-gestational-age fetuses. Am J Obstet Gynecol 1989;160:854-60; discussion 60-2.
83. Schild RL, Sachs C, Fimmers R, Gembruch U, Hansmann M. Sex-specific fetal weight prediction by ultrasound. Ultrasound Obstet Gynecol 2004;23:30-5.
84. Combs CA, Jaekle RK, Rosenn B, Pope M, Miodovnik M, Siddiqi TA. Sonographic estimation of fetal weight based on a model of fetal volume. Obstet Gynecol 1993;82:365-70.
85. Dudley NJ. A systematic review of the ultrasound estimation of fetal weight. Ultrasound Obstet Gynecol 2005;25:80-9.

86. Anderson NG, Jolley IJ, Wells JE. Sonographic estimation of fetal weight: comparison of bias, precision and consistency using 12 different formulae. Ultrasound Obstet Gynecol 2007;30:173-9.
87. Kurmanavicius J, Burkhardt T, Wisser J, Huch R. Ultrasonographic fetal weight estimation: accuracy of formulas and accuracy of examiners by birth weight from 500 to 5000 g. J Perinat Med 2004;32:155-61.
88. Owen P, Donnet ML, Ogston SA, Christie AD, Howie PW, Patel NB. Standards for ultrasound fetal growth velocity. Br J Obstet Gynaecol 1996;103:60-9.
89. Deter RL, Harrist RB, Hadlock FP, Poindexter AN. Longitudinal studies of fetal growth with the use of dynamic image ultrasonography. Am J Obstet Gynecol 1982;143:545-54.
90. Gallivan S, Robson SC, Chang TC, Vaughan J, Spencer JA. An investigation of fetal growth using serial ultrasound data. Ultrasound Obstet Gynecol 1993;3:109-14.
91. Hadlock FP, Harrist RB, Martinez-Poyer J. In utero analysis of fetal growth: a sonographic weight standard. Radiology 1991;181:129-33.
92. Johnsen SL, Rasmussen S, Wilsgaard T, Sollien R, Kiserud T. Longitudinal reference ranges for estimated fetal weight. Acta Obstet Gynecol Scand 2006;85:286-97.
93. Mongelli M, Gardosi J. Longitudinal study of fetal growth in subgroups of a low-risk population. Ultrasound Obstet Gynecol 1995;6:340-4.
94. Schwarzler P, Bland JM, Holden D, Campbell S, Ville Y. Sex-specific antenatal reference growth charts for uncomplicated singleton pregnancies at 15-40 weeks of gestation. Ultrasound Obstet Gynecol 2004;23:23-9.
95. Royston P, Altman DG. Design and analysis of longitudinal studies of fetal size. Ultrasound Obstet Gynecol 1995;6:307-12.
96. Cooke RW. Conventional birth weight standards obscure fetal growth restriction in preterm infants. Arch Dis Child Fetal Neonatal Ed 2007;92:F189-92.
97. Jansson N, Nilsfelt A, Gellerstedt M, et al. Maternal hormones linking maternal body mass index and dietary intake to birth weight. Am J Clin Nutr 2008;87:1743-9.
98. Nohr EA, Vaeth M, Baker JL, Sorensen T, Olsen J, Rasmussen KM. Combined associations of prepregnancy body mass index and gestational weight gain with the outcome of pregnancy. Am J Clin Nutr 2008;87:1750-9.
99. Lackman F, Capewell V, Richardson B, daSilva O, Gagnon R. The risks of spontaneous preterm delivery and perinatal mortality in relation to size at birth according to fetal versus neonatal growth standards. Am J Obstet Gynecol 2001;184:946-53.
100. Zaw W, Gagnon R, da Silva O. The risks of adverse neonatal outcome among preterm small for gestational age infants according to neonatal versus fetal growth standards. Pediatrics 2003;111:1273-7.
101. Bernstein IM, Mohs G, Rucquoi M, Badger GJ. Case for hybrid "fetal growth curves": a population-based estimation of normal fetal size across gestational age. J Matern Fetal Med 1996;5:124-7.
102. Fry AG, Bernstein IM, Badger GJ. Comparison of fetal growth estimates based on birth weight and ultrasound references. J Matern Fetal Neonatal Med 2002;12:247-52.
103. Mongelli M, Biswas A. A fetal growth standard derived from multiple modalities. Early Hum Dev 2001;60:171-7.
104. Gardosi J, Chang A, Kalyan B, Sahota D, Symonds EM. Customised antenatal growth charts. Lancet 1992;339:283-7.
105. Gardosi J, Mongelli M, Wilcox M, Chang A. An adjustable fetal weight standard. Ultrasound Obstet Gynecol 1995;6:168-74.

106. Clausson B, Gardosi J, Francis A, Cnattingius S. Perinatal outcome in SGA births defined by customised versus population-based birthweight standards. BJOG 2001;108:830-4.
107. Mongelli M, Gardosi J. Reduction of false-positive diagnosis of fetal growth restriction by application of customized fetal growth standards. Obstet Gynecol 1996;88:844-8.
108. de Jong CL, Gardosi J, Dekker GA, Colenbrander GJ, van Geijn HP. Application of a customised birthweight standard in the assessment of perinatal outcome in a high risk population. British Journal of Obstetrics & Gynaecology 1998;105:531-5.
109. Lyon V, Howatson A, Khan KS, Owen P. Unadjusted and customised weight centiles in the identification of growth restriction among stillborn infants. BJOG 2004;111:1460-3.
110. Owen P, Ogah J, Bachmann LM, Khan KS. Prediction of intrauterine growth restriction with customised estimated fetal weight centiles. BJOG 2003;110:411-5.
111. McCowan LM, Harding JE, Stewart AW. Customized birthweight centiles predict SGA pregnancies with perinatal morbidity. BJOG 2005;112:1026-33.
112. Sanderson DA, Wilcox MA, Johnson IR. The individualised birthweight ratio: a new method of identifying intrauterine growth retardation. Br J Obstet Gynaecol 1994;101:310-4.
113. Sciscione AC, Gorman R, Callan NA. Adjustment of birth weight standards for maternal and infant characteristics improves the prediction of outcome in the small-for-gestational-age infant. Am J Obstet Gynecol 1996;175:544-7.
114. Ego A, Subtil D, Grange G, et al. Customized versus population-based birth weight standards for identifying growth restricted infants: a French multicenter study. Am J Obstet Gynecol 2006;194:1042-9.
115. Royal College of Obstetricians and Gynaecologists. The Investigation and Management of the Small-for-Gestational-Age Fetus (Guideline 31); 2002.
116. Resnik R. One size does not fit all. Am J Obstet Gynecol 2007;197:221-2.
117. Mongelli M, Figueras F, Francis A, Gardosi J. A customized birthweight centile calculator developed for an Australian population. Aust N Z J Obstet Gynaecol 2007;47:128-31.
118. McCowan L, Stewart AW, Francis A, Gardosi J. A customised birthweight centile calculator developed for a New Zealand population. Aust N Z J Obstet Gynaecol 2004;44:428-31.
119. Bukowski R, Uchida T, Smith GC, et al. Individualized norms of optimal fetal growth: fetal growth potential. Obstet Gynecol 2008;111:1065-76.
120. Chang TC, Robson SC, Spencer JA, Gallivan S. Prediction of perinatal morbidity at term in small fetuses: comparison of fetal growth and Doppler ultrasound. Br J Obstet Gynaecol 1994;101:422-7.
121. Deter RL, Rossavik IK. A simplified method for determining individual growth curve standards. Obstet Gynecol 1987;70:801-6.
122. Deter RL, Rossavik IK, Harrist RB, Hadlock FP. Mathematic modeling of fetal growth: development of individual growth curve standards. Obstet Gynecol 1986;68:156-61.
123. Smith-Bindman R, Chu PW, Ecker JL, Feldstein VA, Filly RA, Bacchetti P. US evaluation of fetal growth: prediction of neonatal outcomes. Radiology 2002;223:153-61.
124. de Jong CL, Francis A, van Geijn HP, Gardosi J. Fetal growth rate and adverse perinatal events. Ultrasound Obstet Gynecol 1999;13:86-9.
125. Guihard-Costa AM, Droulle P, Larroche JC. Growth velocity of the biparietal diameter, abdominal transverse diameter and femur length in the fetal period. Early Hum Dev 1991;27:93-102.
126. Salomon LJ, Bernard JP, Duyme M, Ville Y. Predicting late-onset growth abnormalities using growth velocity between trimesters. J Matern Fetal Neonatal Med 2005;17:193-7.

127. Ariyuki Y, Hata T, Kitao M. Evaluation of perinatal outcome using individualized growth assessment: comparison with conventional methods. Pediatrics 1995;96:36-42.
128. Owen P, Harrold AJ, Farrell T. Fetal size and growth velocity in the prediction of intrapartum caesarean section for fetal distress. Br J Obstet Gynaecol 1997;104:445-9.
129. Owen P, Khan KS. Fetal growth velocity in the prediction of intrauterine growth retardation in a low risk population. Br J Obstet Gynaecol 1998;105:536-40.
130. Hemachandra AH, Klebanoff MA. Use of serial ultrasound to identify periods of fetal growth restriction in relation to neonatal anthropometry. Am J Hum Biol 2006;18:791-7.
131. Williams KP, Nwebube N. Abdominal circumference: a single measurement versus growth rate in the prediction of intrapartum Cesarean section for fetal distress. Ultrasound Obstet Gynecol 2001;17:493-5.
132. Hooper PM, Mayes DC, Demianczuk NN. A model for foetal growth and diagnosis of intrauterine growth restriction. Stat Med 2002;21:95-112.
133. Pan H, Goldstein H. Multi-level models for longitudinal growth norms. Stat Med 1997;16:2665-78.
134. Royston P. Calculation of unconditional and conditional reference intervals for foetal size and growth from longitudinal measurements. Stat Med 1995;14:1417-36.
135. Owen P, Ogston S. Conditional centiles for the quantification of fetal growth. Ultrasound Obstet Gynecol 1998;11:110-7.
136. Koops BL, Morgan LJ, Battaglia FC. Neonatal mortality risk in relation to birth weight and gestational age: update. J Pediatr 1982;101:969-77.
137. Kornfeld D, Cnattingius S, Ekbom A. Pregnancy outcomes in women with inflammatory bowel disease--a population-based cohort study. Am J Obstet Gynecol 1997;177:942-6.
138. Luo ZC, Simonet F, An N, Bao FY, Audibert F, Fraser WD. Effect on neonatal outcomes in gestational hypertension in twin compared with singleton pregnancies. Obstet Gynecol 2006;108:1138-44.
139. Xiong X, Demianczuk NN, Buekens P, Saunders LD. Association of preeclampsia with high birth weight for age. Am J Obstet Gynecol 2000;183:148-55.
140. Hernan MAS, E.F.; Hernandez-Diaz, S. Prenatal exposures, gestational age, birth weight, and mortality: an application of causal diagrams. In: 17th Annual Meeting of the Society for Pediatric and Perinatal Epidemiologic Research; 2004: 38; Salt Lake City, Utah: Society for Pediatric and Perinatal Epidemiologic Research; 2004: 38.
141. Platt RW, Joseph KS, Ananth CV, Grondines J, Abrahamowicz M, Kramer MS. A proportional hazards model with time-dependent covariates and time-varying effects for analysis of fetal and infant death. Am J Epidemiol 2004;160:199-206.
142. Savitz DA, Hertz-Picciotto I, Poole C, Olshan AF. Epidemiologic measures of the course and outcome of pregnancy. Epidemiol Rev 2002;24:91-101.
143. Raghunathan TE. What do we do with missing data? Some options for analysis of incomplete data. Annu Rev Public Health 2004;25:99-117.
144. Rubin DB. Inference and missing data. Biometrika 1976;63:581-92.
145. Schafer JL, Graham JW. Missing data: our view of the state of the art. Psychol Methods 2002;7:147-77.
146. Greenland S, Finkle WD. A critical look at methods for handling missing covariates in epidemiologic regression analyses. Am J Epidemiol 1995;142:1255-64.
147. Demissie K, Rhoads GG, Ananth CV, et al. Trends in preterm birth and neonatal mortality among blacks and whites in the United States from 1989 to 1997. Am J Epidemiol 2001;154:307-15.

148. Joseph KS, Kramer MS, Marcoux S, et al. Determinants of preterm birth rates in Canada from 1981 through 1983 and from 1992 through 1994. N Engl J Med 1998;339:1434-9.
149. Min SJ, Luke B, Gillespie B, et al. Birth weight references for twins. Am J Obstet Gynecol 2000;182:1250-7.
150. Odegard RA, Vatten LJ, Nilsen ST, Salvesen KA, Austgulen R. Preeclampsia and fetal growth. Obstet Gynecol 2000;96:950-5.
151. Shiono PH, Klebanoff MA, Rhoads GG. Smoking and drinking during pregnancy. Their effects on preterm birth. JAMA 1986;255:82-4.
152. Wilcox AJ. Birth weight and perinatal mortality: the effect of maternal smoking. Am J Epidemiol 1993;137:1098-104.
153. Joseph KS. Incidence-based measures of birth, growth restriction, and death can free perinatal epidemiology from erroneous concepts of risk. J Clin Epidemiol 2004;57:889-97.
154. Salihu HM, Sharma PP, Aliyu MH, et al. Is small for gestational age a marker of future fetal survival in utero? Obstet Gynecol 2006;107:851-6.
155. Robins JM, Rotniztky A, Zhao LP. Estimation of regression coefficients when some regressors are not always observed. Journal of the American Statistical Association 1994;89:846-66.
156. Buekens P, Klebanoff M. Preterm birth research: from disillusion to the search for new mechanisms. Paediatr Perinat Epidemiol 2001;15 Suppl 2:159-61.
157. Gelman A, Carlin JB, Stern HS, Rubin DB. Bayesian data analysis. 2nd ed. London, United Kingdom: Chapman & Hall/CRC; 2003.
158. The National Standard for Normal Fetal Growth. 2007. (Accessed September 21, 2008, at http://www.nichd.nih.gov/about/org/despr/studies/preg/normfetalgrowth.cfm.)
159. Cnattingius S, Ericson A, Gunnarskog J, Kallen B. A quality study of a medical birth registry. Scand J Soc Med 1990;18:143-8.
160. Socialstyrelsens Epidemiologiska Centrum. Utvärdering av det svenska Medicinska födelseregistret. Article 2002-112-4. http://www.socialstyrelsen.se/Publicerat/2002/2023/2002-112-4.htm. Stockholm, Sweden; 2002.
161. Yudkin PL, Wood L, Redman CW. Risk of unexplained stillbirth at different gestational ages. Lancet 1987;1:1192-4.
162. Groom KM, Poppe KK, North RA, McCowan LM. Small-for-gestational-age infants classified by customized or population birthweight centiles: impact of gestational age at delivery. Am J Obstet Gynecol 2007;197:239 e1-5.
163. Hutcheon JA, Platt RW. The missing data problem in birth weight percentiles and thresholds for "small-for-gestational-age". Am J Epidemiol 2008;167:786-92.
164. Chard T. Does the fetus lose weight in utero following fetal death: a study in preterm infants. BJOG 2001;108:1113-5.
165. Beaty TH, Skjaerven R, Breazeale DR, Liang KY. Analyzing sibship correlations in birth weight using large sibships from Norway. Genet Epidemiol 1997;14:423-33.
166. Basso O, Frydenberg M, Olsen SF, Olsen J. Two definitions of "small size at birth" as predictors of motor development at six months. Epidemiology 2005;16:657-63.
167. Battaglia FC, Frazier TM, Hellegers AE. Birthweight, gestational age, and pregnancy outcome, with special reference to high birthweight-low gestational age infants. Pediatrics 1966;37:417-22.

168. Deter RL, Harrist RB, Hill RM. Neonatal growth assessment score: a new apporach to the detection of intrauterine growth retardation in the newborn. Am J Obstet Gynecol 1990;162:1030-6.
169. Owen P, Burton K, Ogston S, Khan KS, Howie PW. Using conditional and unconditional standard deviation scores of fetal abdominal area measurements in the prediction of intrauterine growth restriction. . Ultrasound Obstet Gynecol 2000;16:439-44.
170. Hutcheon JA, Zhang X, Cnattingius S, Kramer MS, Platt RW. Customised birthweight percentiles: does adjusting for maternal characteristics matter? BJOG 2008;115:1397-404.
171. Ross S. Composite outcomes in randomized clinical trials: arguments for and against. Am J Obstet Gynecol 2007;196:119 e1-6.
172. Frisen M, de Mare J. Optimal surveillance. Biometrika 1991;78:271-80.
173. Harrell FE. Regression Modeling Strategies with Applications to Linear Models, Logistic Regression and Survival Analysis. New York, USA.: Springer-Verlag; 2001.
174. Silverwood RJ, Cole TJ. Statistical methods for constructing gestational age-related reference intervals and centile charts for fetal size. Ultrasound Obstet Gynecol 2007;29:6-13.
175. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 1982;143:29-36.
176. Bakketeig LS, Jacobsen G, Hoffman HJ, et al. Pre-pregnancy risk factors of small-for-gestational age births among parous women in Scandinavia. Acta Obstet Gynecol Scand 1993;72:273-9.
177. Goldenberg RL, Cliver SP, Neggers Y, et al. The relationship between maternal characteristics and fetal and neonatal anthropometric measurements in women delivering at term: a summary. Acta Obstet Gynecol Scand Suppl 1997;165:8-13.
178. Apgar V. A proposal for a new method of evaluation of the newborn infant. Curr Res Anesth Analg 1953;32:260-7.
179. The Swedish Centre for Epidemiology. The Swedish Medical Birth Register- A summary of content and quality. Article 2003-112-3. Stockholm, Sweden; 2003.

Appendix A. List of Tables & Figures

Tables

Figures

Appendix B. List of abbreviations

AGA- Appropriate for gestational age

AUC- Area under the curve

BMI- Body mass index

CI- Confidence interval

EFW- Estimated fetal weight

IUGR- Intrauterine growth restriction

LBW- Low birth weight

LMP- Last menstrual period

MAR- missing at random

MCAR- missing completely at random

MNAR- missing not at random

MOND- McGill Obstetric and Neonatal Database

NICU- Neonatal intensive care unit

OR- Odds ratio

ROC-Receiver operating characteristic

RR- Relative risk

RVH- Royal Victoria Hospital

SGA- Small for gestational age

SD- Standard deviation

UK- United Kingdom

Appendix C. Description of study populations

E.1 Royal Victoria Hospital cohort

The Royal Victoria Hospital (RVH) is a McGill University tertiary care teaching hospital in Montreal, Canada that serves a multi-ethnic population of predominantly Caucasian, Middle Eastern, and Asian origin. Since 1978, the obstetrical and neonatal medical charts of all pregnancies delivered at the Royal Victoria have been entered into a database known as the McGill Obstetric and Neonatal Database (MOND), developed and maintained by the late Dr. Robert Usher. The MOND is a comprehensive, quality controlled clinical database that contains over 250 variables on maternal characteristics, obstetrical complications, clinical management, and neonatal outcomes. For the time periods included in this thesis (1996- 2006), all database entry was performed by a single data archivist. The original database, known as MOND1, was used from 1978 until March 31th, 2001, when it was replaced by an updated and expanded version, known as MOND2. For administrative reasons, MOND data were not available from March 31st 1997- April 1st 1998 and March 31st 2000- April 1st 2001.

Using the mother's hospital case number as a unique identifier, the MOND database was linked with the electronic ultrasound records of the Royal Victoria Hospital's obstetrical ultrasound department. For women with more than one delivery at the Royal Victoria, ultrasounds were matched based on plausibility of dates of ultrasound and delivery. The key variables of the combined databases used in this thesis are described below:

Gestational age

Gestational age at birth in days was obtained from the MOND. In the majority of pregnancies (82%), gestational age was calculated from the first day of the last normal menstrual period (LMP), confirmed by early ultrasound (<20 weeks). If the gestational age calculated with LMP differed from the estimate of gestational age based on early ultrasound dating by more than 10 days, the latter was used (11% of pregnancies). Since a threshold of 7 days discrepancy was used during the MOND 1 period, gestational ages from this period were re-calculated to be consistent with the 10-day threshold currently used. In 6 percent of pregnancies, gestational age was based on LMP in the absence of confirmation by early ultrasound. Gestational age was based on early ultrasound in the absence of information on LMP in less than 1 percent of pregnancies.

Birth weight

Birth weight was obtained from the MOND. Newborn infants are weighed nude on an electronic scale by the attending physician or resident, and birth weight is recorded to the nearest gram. There are no missing birth weights for the time periods included in this thesis.

Estimated fetal weight (EFW)

Estimated fetal weight in grams was obtained from the RVH obstetrical ultrasound database. Between 1995 and 2006, the RVH had a policy of performing a routine ultrasound examination at 32- weeks (which in practice, was done at 32-33 weeks), including an estimation of fetal weight. Fetal weight is estimated through a formula that combines biometric measurements of the abdomen, femur, and head(described below). The ultrasound machines at the RVH are programmed to calculate EFW using the formula of Hadlock,[79] which relates fetal weight exponentially to measurements of the abdominal circumference(AC), femur length(FL), biparietal diameter(BPD), and head circumference(HC) ($EFW=10^x$ where $x=1.3596 + (0.0424 \times AC) + (0.174 \times FL) + (0.0064 \times HC) + (0.0061 \times AC \times BPD) - (0.00386 \times AC \times FL)$). In this thesis, however, estimates of fetal weight were calculated using Combs' formula,[84] which is based on a volumetric approximation of fetal weight (a 2-compartment model based on the volume of the head and trunk). The volumetric approach of Combs' formula makes it more robust on the extreme ends of the birth weight distribution than regression-based approaches. It has been found to have a lower random and systematic error than other formula to estimate fetal weight (DU DLEY). Combs' formula is based on abdominal circumference, femur length, and head circumference measurements ($EFW=(0.23718 \times AC^2 \times FL) + (0.03312 \times HC^3)$, where EFW is in grams and AC, FL, HC are in cm).

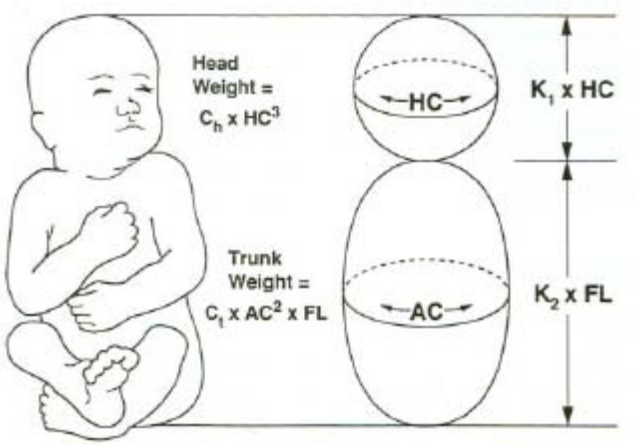


Figure 1. Two-compartment model of fetal weight. AC = abdominal circumference; FL = femur length; HC = head circumference; K_x, C_x = constants. Algebraically, $C_1 = DK_2/6\pi$ and $C_h = DK_1/6\pi$, where D = density.

Source: Combs CA, Jaekle RK, Rosenn B, et al. Sonographic estimation of fetal weight based on a model of fetal volume. Obstet Gynecol 1993;82:365-70. Reproduced with permission, copyright Lippincott, Williams, & Wilkins.

Abdominal circumference. measured in millimetres (mm) to the nearest tenth of a mm using standard techniques.

Femur length: measured in millimetres (mm) to the nearest tenth of a mm using standard techniques.

Head circumference. measured in millimetres (mm) to the nearest tenth of a mm using standard techniques.

Sex

Sex is coded in the MOND as either male, female or ambiguous. There were no missing or ambiguous values in the MOND for the time periods included in this thesis.

Perinatal mortality

Perinatal mortality was based on stillbirths or in-hospital neonatal deaths following initial admission (birth) or re-admission recorded in the MOND. Births under 500 grams (live or stillborn) are not included in the MOND database. As with other research studies performed with MOND data, linkage to vital statistics was not made to establish neonatal deaths that occurred outside of hospital.

5 minute Apgar score

The Apgar score at 5 minutes after birth was obtained from the MOND. The Apgar score reflects an evaluation of the newborn's health by the attending physician or resident, and is a sum total of 5 different measures of newborn status: skin colour, heart rate, reflex irritability, muscle tone, and breathing.[178] Each measure is assigned a score of 0,1, or 2, with 0 indicating "absent" and 2 indicating "present". The final Apgar score therefore ranges between 0 and 10, with 10 indicating the best possible state.

Cord pH

Umbilical cord blood pH levels were obtained from the MOND. Samples were analyzed at the Royal Victoria Hospital Laboratory, with possible values ranging from 6.00 to 8.00. Cord pH values were missing in 13% of the MOND population, with missing values being more likely to be from earlier time periods in the dataset (28% pre-2001 versus 6% post-2001). Values were not missing completely at random: in particular, women with missing values were more likely to have delivered vaginally (Caesarean section rate of 7% among those with missing values versus 26% among those with available values). Differences in birth weight and gestational age at birth between those with cord pH values and missing cord pH values were statistically significant, but not clinically important (infants with missing values were an average of 51 grams smaller and less than 2 days younger). There were no differences in birthweight ratio between the two groups, indicating that differences in weight between the two groups were due to differences in gestational age at birth.

Other adverse outcome variables (neonatal convulsions, renal failure due to asphyxia, need for intubation due to asphyxia, pulmonary hypertension due to asphyxia, and cardiac insufficiency due to asphyxia) were based on MOND data extracted from medical charts records. Neonatal records were reviewed and coded by a single neonatologist during the study period (Dr. Robert Usher) to ensure consistency and accuracy of the database entries.

E.2 Swedish Medical Birth Register

Since 1973, the Swedish Medical Birth Register has recorded data on the antenatal care and pediatric examination of births in Sweden.[159] The SMBR is a quality-controlled database and includes records for 98-99 percent of births. All women and births in Sweden are assigned an 8-digit Personal Identification Number (PIN), which allows linkage of the Medical Birth Register to other national databases such as the Cause-of-Death Register of the Swedish Centre for Epidemiology and Statistics Sweden's Birth Register. Descriptions of the variables used in this thesis are excerpted from "The Swedish Medical Birth Register- a summary of content and quality" published by the Swedish Centre for Epidemiology.[179]

Birth weight

Information on birth weight is lacking for 0.32 per cent of all infants, the rate varying between 0.04 and 1.95 per cent (1994). Some absurd values are given, e.g., less than 300g; but the majority of weights are plausible.

Pregnancy duration

In order to estimate pregnancy duration, a number of variables are used:

- date of last menstrual period (LMP)
- estimated day of delivery (from LMP, possibly modified from clinical investigations)
- corrected estimated day of delivery (from second-trimester ultrasound)
- pregnancy duration as stated in paediatric record

All dates are sensitive to mis-representation, and pregnancy duration estimates may therefore be incorrect. The paediatric estimate (in completed weeks) may also be incorrect; a mistake in one digit may cause a large error. To get the "best possible" estimate of pregnancy duration, the following hierarchic rules are applied (but the basic data listed above are kept in the register):

1. Pregnancy duration from corrected estimated day of delivery is supported by information in the paediatric record and is then kept. For 1998, this was true for 81.8 per cent of pregnancies.

2. Pregnancy duration from estimated day of delivery is supported by information in the paediatric record and is then kept; true for 8.0 per cent (usually because corrected dates are lacking).
3. The only available information is the paediatric estimate which is then kept; true for 0.65 per cent.
4. Pregnancy duration estimated from LMP date is supported by information in the paediatric record and is then kept; true for 7.5 per cent.
5. Pregnancy duration from corrected estimated date of delivery is supported by pregnancy duration from estimated date of delivery; the former is kept; true for 0.9 per cent.
6. Pregnancy duration from corrected estimated date of delivery is supported by pregnancy duration from LMP; the former is kept; true for 0.1 per cent.
7. Pregnancy duration from corrected estimated date of delivery is the only available information and is kept; true for 0.1 per cent.
8. Pregnancy duration from estimated date of delivery is the only available information and is kept; true for 0.01 per cent.
9. Pregnancy duration estimated from LMP is the only available information and is kept; true for 0.02 per cent.
10. Pregnancy duration from estimated date of delivery and from LMP agree; the former is kept; true for 0.02 per cent.
11. If information on infant sex or birth weight is missing, if it is a multiple birth, or a delivery of a severely malformed infant, an estimate of pregnancy duration is made according to the following hierarchy:

    estimated from corrected expected date of delivery
    estimated from expected date of delivery
    estimated from LMP date
    paediatric record information

    This is true for 0.05 per cent of cases.
12. In all cases where disparate information exists (and none of the above listed combinations are fulfilled), the estimated pregnancy duration which gives the smallest deviation of birth weight from the normal birth weight/pregnancy duration distribution is kept; true for 0.8 per cent.

Thus, in the majority of cases (97 per cent) pregnancy duration is based on the agreement between a calculated pregnancy duration with that stated on the paediatric record. Among the remaining cases, rather large discrepancies can be found (>5 weeks).

Live birth and stillbirth

Stillbirths can be identified in two ways: a mark in a check box or from the infant's personal identification number. There is a slight discrepancy between stillbirths in the Medical Birth Register and in the register of Statistics Sweden. The most frequent difference is for stillbirths born before 28 full weeks of pregnancy. According to Swedish law, such births are to be regarded as late abortions, not stillbirths, and should not be included in statistics on delivered infants. In some cases, such a foetus has been recorded in both registers, in other cases not in that of Statistics Sweden. Of the infants recorded as live-born in the Medical Birth Register, 0.2 percent are not included in the Statistics Sweden data. The same is true for 2.6 per cent of stillbirths.

There are 18 infants marked as stillbirths in the Medical Birth Registry but with a personal identification number indicating live birth - all 18 are wrongly marked in the registry due to a misrepresentation (5 in 1986, 8 in 1990, 5 in 1991).

Infant sex

The sex of the newborn is noted in the paediatric record, and is also shown by the personal identification number. Among the more than two million infants born, sex information is missing for 200 of those with an identification number (130 boys, 70 girls), and 164 infants had no identification number. For 0.05 per cent of the infants with identification numbers, the sex noted in the paediatric record does not agree with that shown by the identification number.

Mother's pre-pregnancy weight

These data have been compiled from 1983 onwards. Pre-pregnancy weight (in effect, weight at first antenatal-care visit) data are available for 70 per cent of the women. For the years 1990 - 1991, practically no data are available; for the other years, the percentage varies between 73-88 per cent. Information on weight gain during pregnancy is available in about 60 per cent of

women, ranging from 29.86 per cent for various years. For technical reasons, the figures are uncertain for the period 1983 -1990...Height is known for about 80 per cent of the mothers, and only relatively few values are invalid. From these data, it is possible to calculate the body mass index (BMI) for about 65 per cent of the women with a reasonable accuracy. (Data are missing for two years; otherwise, the rate varies between 65.85 per cent.)

Parity

There are two sources for parity: One is based on the number of reported previous stillbirths and live births (+1), the other on data from Statistics Sweden. When the two sources are compared, major discrepancies are found. These data can also be compared with the number of previously registered infants in the Medical Birth Register (estimated parity). Analysis shows that for singleton births, the parity data from Statistics Sweden agrees with the estimated parity in 98 per cent of cases. For women born in Sweden, the parity information based on reported previous births in the Medical Birth Register agrees with the estimated parity in 91 per cent of cases; but for the period from 1982-1989, the level of agreement is only 85-89 per cent, and for 1990 only 54 per cent. For women not born in Sweden, the difference between the parity data from Statistics Sweden and the estimated parity is larger: the level of agreement is only 91 per cent. Agreement with the Medical Birth Register, is also 91 per cent (similar to the level for women born in Sweden). Thus, the best parity estimate is that obtained from Statistics Sweden. But especially for immigrant women, some errors exist due to the fact that registration of births which take place outside Sweden is incomplete.

Appendix D. Calculation of conditional percentiles

Conditional fetal growth percentiles are calculated through the steps described below. A worked example of the construction of conditional percentiles is also provided for illustration.

STEP 1. Mixed model

A mixed model is first built to describe the longitudinal estimated fetal weight measurements as a function of gestational age. This produces the necessary estimates of fetal weight population averages of estimated fetal weights at each gestational age, as well as of the variability between- and within- fetuses in weight.

The following formula describes a basic mixed model whereby both the intercept ($_0$) and slope ($_{GA}$) are allowed to vary by fetus:

(1) $\quad _0$

where:

'i' denotes the 'ith' fetus, 'j' denotes the 'jth' measurement occasion, EFW denotes the response variable, "Estimated Fetal Weight", and denotes the within-fetus variability in fetal weights

X is the independent variable of gestational age

$_0 \quad _0$; $_i$ being the random effect (latent variable) at the level of the fetus to allow each fetus to have its own intercept.

$( ) \quad ( )$; $_{GA(i)}$ again being the random effect at the level of the fetus, here, allowing each fetus to have its own slope (ie. taking into account that fetuses will grow at different rates during pregnancy).

$\text{var} \quad _0 \quad ^2_0$

$\text{var} \quad ^2$

$\text{cov}(\ _0 , \quad) \quad _0$

The variance of EFW for fetus *i* at time *j* (the unconditional variance) is calculated as:

(2) $\text{var}(\quad) \quad {}^{2}_{0} \quad {}^{2} \quad {}^{2} \quad 2 \quad {}_{0,} \quad {}^{2}$

STEP 2. Conditional means and variances

The conditional mean EFW at time 2, given the fetus' EFW at time 1 is calculated as:

(3) $(\quad {}_{2}| \quad {}_{1}) \quad {}_{2|1} \quad {}_{2} \quad (\quad {}_{1} \quad {}_{1}) \frac{{}_{12}}{{}^{2}_{1}}$

where:

$${}_{12} \quad \text{cov}(\quad {}_{1}, \quad {}_{2})$$

$$\text{cov}({}_{0} \quad {}_{1} \quad {}_{1}, {}_{0} \quad {}_{2} \quad {}_{2})$$

$${}^{2}_{0} \quad {}_{0,} \quad ({}_{1} \quad {}_{2}) \quad {}_{1} \quad {}_{2} \quad {}^{2}$$

The conditional variance of EFW_2 given EFW_1 is:

(3) $(\quad {}_{2}| \quad {}_{1}) \quad {}^{2}_{2|1} \quad {}^{2}_{2} \quad \frac{{}^{2}_{12}}{{}^{2}_{1}}$

STEP 3. Conditional reference intervals

With conditional means and variances established, the 80% conditional reference interval of EFW_2 given EFW_1 is calculated as:

(4) $\exp(\quad {}_{2|1} \quad 1.28 \quad {}_{2|1})$

Worked Example

The above formulae represent an overly simplistic approach to the calculation of conditional fetal growth percentiles. In practice, additional complexities such as the non-linear relationship between gestational age and fetal weight, and the influence of fetal sex on weight would be

taken into account. To illustrate a more realistic calculation of conditional percentiles, a worked example is provided using data from the Scandinavian portion of the 1986-88 Study of Successive Small-for-Gestational Age Births project funded by the US National Institute of Child Health and Human Development (NICHD).

Population:

The Successive SGA Births Study collected longitudinal ultrasounds measurements from nearly 2000 pregnant women of parity 1 or 2. These women consisted of both a 10% random sample (n=561) of all women eligible for the study (n= 5722), as well as all eligible women at high risk of a small-for-gestational-age (SGA) birth (n=1384). The risk factors for SGA were considered to be: 1) prior low birth weight birth 2) maternal cigarette smoking at conception 3) low pre-pregnancy weight 4) prior perinatal death 5) chronic maternal disease (chronic renal disease, essential hypertension, or heart disease). An exception to 100% inclusion of women who met high risk criteria was smokers, from whom a 50% sample was included. The participation rate in this study was 80%, with reasons for non-participation largely due to long travel times needed to attend additional study prenatal clinic visits. Further details on the study are available elsewhere.[176]

Ultrasound measurements:

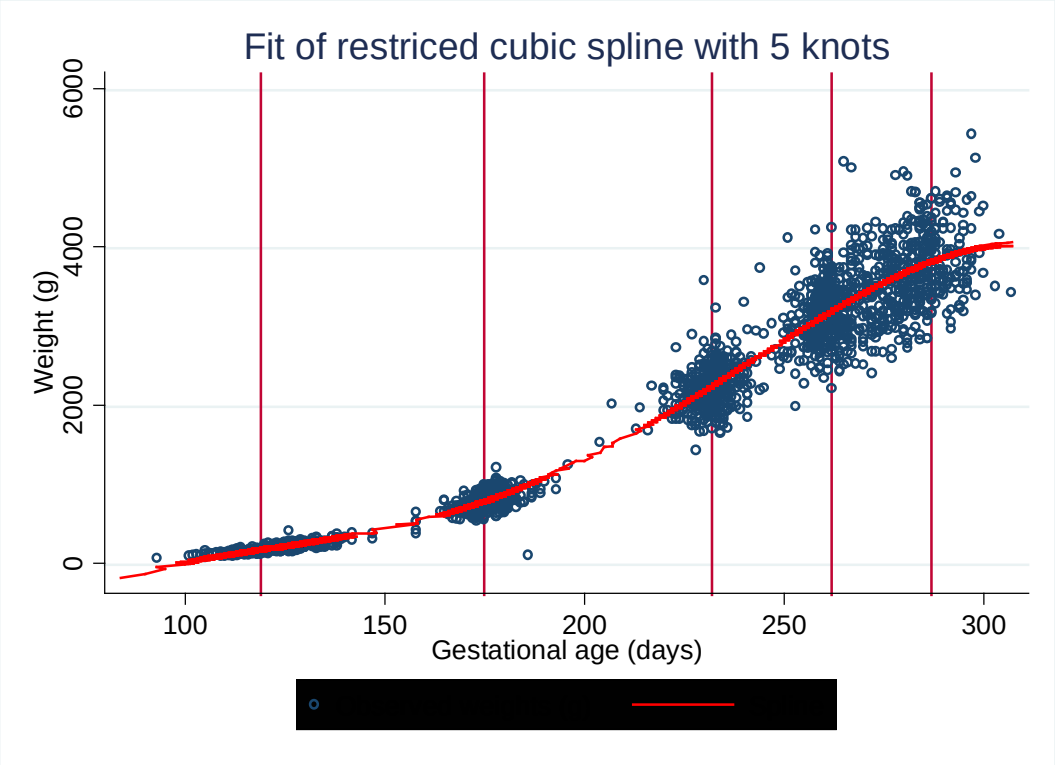
Fetal biparietal diameter, femur length, and abdominal diameter were measured at 17, 25, 33, and 37 weeks of pregnancy, and were combined to estimate fetal weight using the formula of Hadlock.[79] Removing women who did not have at least 1 valid estimate of fetal weight in addition to birth weight left n=449 women with 2138 weight measurements (average of 4.8 per woman) in the 10% random sample to estimate the coefficients for the conditional percentiles.

Coefficients for conditional percentiles:

The mixed model to estimate the coefficients for conditional percentiles is built with the following considerations:

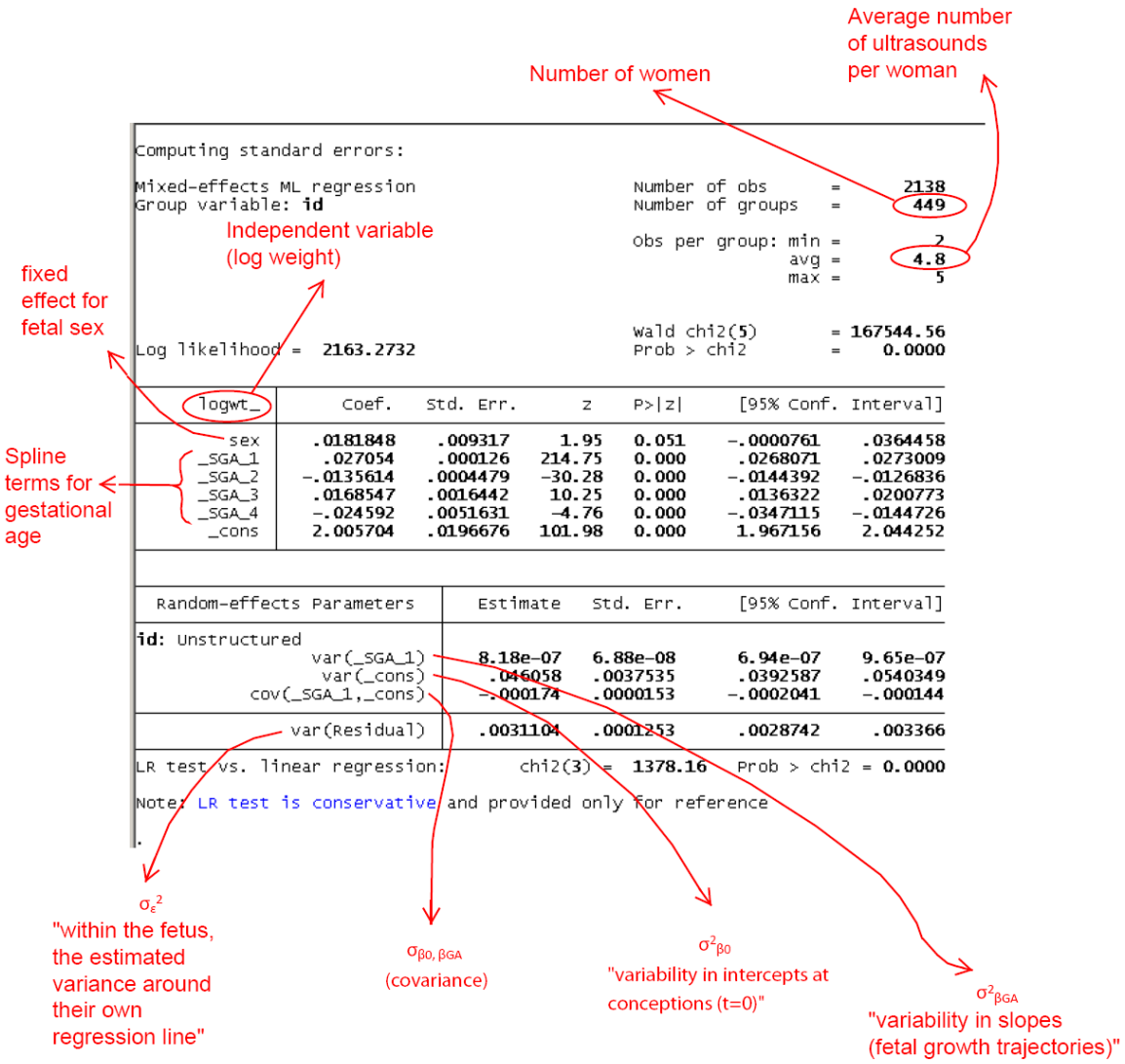
1) Estimated fetal weight is log-transformed (natural logarithm) in order to ensure that the assumption of homoscedascity of model residuals is not violated.
2) Sex is included as a fixed-effect covariate in the model on substantive grounds, since there are well-established physiological differences in fetal weights according to sex.

3) Gestational age is modelled as a restricted cubic spline to account for the non-linear relationship of gestational age and estimated fetal weight. Restricted cubic splines are piecewise segments of cubic polynomial functions used to create a continuous, smooth function that better reflects the "shape" of the predictor variable of interest, as shown below. A restricted cubic spline with 5 knots in the default positions (as recommended)[173] provided the best fit.



4) A random effect is included on the linear term of the spline, allowing between-fetus variability in growth curves to be estimated (no assumption is imposed that all fetuses follow similar growth trajectories; fetus-specific trajectories are possible).

The resulting parameters from this mixed model, as produced by STATA 9.0 (Statacorp, College Station TX) are:



Sample calculation of conditional percentiles:

The conditional percentile for a female fetus in this dataset weighing 3360g at birth (aged 270 days (38 weeks)), given that it weighed 2692g at the time of its ultrasound at 234 days (33 weeks) would be calculated as:

1) **Unconditional mean ln(weight) at birth**:
   =2.005704 + 0.0181848*(sex) + 0.027054*(GA Spline term 1) – 0.0135614*(GA Spline term 2) + 0.0168547*(GA Spline term 3) – 0.024592*(Spline term 4)
   =2.005704 + 0.0181848*(0) + 0.027054*(270) – 0.0135614*(121.8647) + 0.0168547*(30.29625) – 0.024592*(1.904252)
   = **8.1214**
   (3365 grams)

2) **Unconditional mean ln(weight) at 234 day ultrasound (32 weeks, 3 days):**
   =2.005704 + 0.0181848*(sex) + 0.027054*(GA Spline term 1) – 0.0135614*(GA Spline term 2) + 0.0168547*(GA Spline term 3) – 0.024592*(Spline term 4)
   =2.005704 + 0.0181848*(0) + 0.027054*(234) – 0.0135614*(53.88588) + 0.0168547*(7.27675) – 0.024592*(0.0002834)
   = 7.7282
   (2272 grams)

3) **Covariance:**
   $\begin{matrix} 2 \\ 0 \end{matrix}$ $\begin{matrix} \\ 0, \end{matrix}$ $(\quad 1 \quad 2)$ $\quad 1 \quad 2$ ^{2}
   = 0.046058 + (-0.000174)*(234+270) + (234*270*(8.18*10E-7))
   = .0100432

4) **Conditional mean ln(weight) at birth given weight at 234 days:**
   $= \quad (\quad 2 \left| \right. \quad 1) \quad 2 \,|1 \quad 2 \quad (\quad 1 \quad 1) \frac{12}{\begin{matrix} 2 \\ 1 \end{matrix}}$
   = 8.1214 + (7.8979253 - 7.7282) x (.0100432/.0125268)
   = 8.257502
   (3856 g)

5) **Conditional variance at birth given weight at 234 days:**
   $= \quad (\quad 2 \left| \right. \quad 1) \quad \begin{matrix} 2 \\ 2\,|1 \end{matrix} \quad \begin{matrix} 2 \\ 2 \end{matrix} \quad \frac{\begin{matrix} 2 \\ 12 \end{matrix}}{\begin{matrix} 2 \\ 1 \end{matrix}}$
   =0.0148406 –(.0100432^2/.0125268)
   = **.00678859**

6) Calculation of 80% coverage limit lower limit (10th conditional percentile):
$= \exp(\quad _{2|1} \quad 1.28 \quad _{2|1})$
= exp(8.257502-1.28*(√(.00678859)))

= **3470 g**

Observed birthweight of 3360g < 10th conditional percentile weight of 3470g, therefore infant is SGA by conditional percentiles.

Appendix E. Certificates of ethical approval

Centre universitaire de santé McGill
McGill University Health Centre
Les meilleurs soins pour la vie
The Best Care for Life

Bureau d'éthique de la recherche
Office of Research Ethics

November 1, 2007

Dr. Robert Platt
4060 Ste Catherine Street West
Room 205
Westmount, Quebec
H3Z 2Z3

Re: "The Predictive Value of Conditional Fetal Growth Centiles"

Dear Dr. Platt:

We are pleased to inform you that the study was found to be within ethical guidelines for conduct at the McGill University Health Centre. Approval for the study protocol (January 2007) was provided via expedited review of the Chair on October 30, 2007 and will be reported to the Research Ethics Board (REB) at its meeting of November 21, 2007. This decision will be entered accordingly into the minutes. At the MUHC, sponsored research activities that require US federal assurance are conducted under Federal Wide Assurance (FWA) 00000840.

All research involving human subjects requires review at a recurring interval and the current study approval is in effect until October 29, 2008. It is the responsibility of the principal investigator to submit an Application for Continuing Review to the REB prior to the expiration of approval to comply with the regulation for continuing review of "at least once per year".

The Research Ethics Boards (REBs) of the McGill University Health Centre are registered REBs working under the published guidelines of the Tri-Council Policy Statement, in compliance with the "Plan d'action ministériel en éthique de la recherche et en intégrité scientifique" (MSSS, 1998) and the Food and Drugs Act (7 June, 2001), acting in conformity with standards set forth in the (US) Code of Federal Regulations governing human subjects research, and functioning in a manner consistent with internationally accepted principles of good clinical practice.

We wish to advise you that this document completely satisfies the requirement for Research Ethics Board Attestation as stipulated by Health Canada.

...2

687, avenue des Pins ouest, S11, Montréal (Québec) H3A 1A1, Tél.: 514-934-1934, Poste 36077/34323, Téléc.: 514-843-1486

The project was assigned MUHC Study Number SDR-07-026 that is required as MUHC reference when communicating about the research. Should any revision to the study, or other unanticipated development occur prior to the next required review, you must advise the REB without delay. Regulation does not permit initiation of a proposed study modification prior to REB approval for the amendment.

We trust this will meet with your complete satisfaction.

Sincerely,

Thonmas Maniatis, MD, MSc, FRCPC
Chair,
SDR Committee

Cc: SDR-07-026

Centre universitaire de santé McGill
McGill University Health Centre

November 5, 2007

Dr. Robert Platt
McGill Epidemiology & Biostatistics
C/O Jennifer Hutcheon
4060, rue Ste Catherine O., Stuioe 205
The Montreal Children's Hospital Research Institute

RE: An evaluation of the predictive value of conditional Fetal growth percentiles.
RVH #071009A.

Dear Dr. Platt:

The impact analysis, the study protocol and worksheet analysis of the above referenced study have been reviewed and it is our understanding that this study has no impact on our hospital resources.

This approval also includes access to review the electronic database for the purpose of this study if required.

If my interpretation of this study's impact is incorrect or requires clarification, please do not hesitate to contact me.

Dr. Françoise P. Chagnon, MDCM, FRCSC, FACS, MHA
Director, Professional Services
McGill University Health Centre

FPC/eh

c.c. Jennifer Hutcheon – study coordinator
D. Pothier/L. Cholette – medical records
S. Lee – research institute

McGill

Faculty of Medicine
3655 Promenade Sir William Osler
Montreal, QC H3G 1Y6

Faculté de médecine
3655, Promenade Sir William Osler
Montréal, QC, H3G 1Y6

Fax/Télécopieur: (514) 398-3595

November 15, 2007

Dr. Robert Platt
Department of Epidemiology and Biostatistics
Montreal Children's Research Institute
4060 Ste. Catherine Ouest, #205
Montreal, Quebec
H3Z 2H3

Dear Dr. Platt,

Thank you for submitting, on behalf of your PhD candidate, Jennifer Hutcheon, the request for review by the IRB of the study proposal entitled "Customized Birth Weight Percentiles: Does Adjusting for Maternal Characteristics Matter?".

As this study involves no more than minimal risk, and in accordance with Article 1.6 of the Canadian Tri-Council Policy Statement of Ethical Conduct for Research Involving Humans and U.S. Title 45 CFR 46, Section 110 (b), paragraph (1), we are pleased to inform you that approval for the study (October 22, 2007) was provided via an expedited review by the Chair on November 15, 2007, valid until **November 2008**. The study proposal will be presented for corroborative approval at the next meeting of the Committee and a certification document will be issued to you at that time.

A review of all research involving human subjects is required on an annual basis in accord with the date of initial approval. The annual review should be submitted at least one month before **November 2008**. Should any modification to the study occur over the next twelve months, please advise IRB appropriately.

Yours sincerely,

Serge Gauthier, M.D.
Chair
Institutional Review Board

cc: Ms. Jennifer Hutcheon
A11-E38-07B

Appendix F. Signed waivers from co-authors

To whom it may concern.

I, the undersigned, give permission for the scholarly use of a manuscript of which I am a co-author in the Doctoral thesis of Jennifer Anne Hutcheon, McGill University Department of Epidemiology, Biostatistics, and Occupational Health:

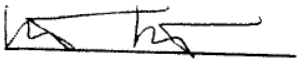
The missing data problem in birth weight percentiles and thresholds for "small-for-gestational-age" (Am J Epidemiol 2008; 167:786-792).

I understand that this manuscript is intended for inclusion in a thesis by manuscript according to the rules published by McGill University at:
http://www.mcgill.ca/gps/current/programs/thesis/guidelines/preparation/

Under these rules, I waive any issues of copyright related to the inclusion of this manuscript in the Doctoral thesis.

Sincerely,

ROBERT PLATT — Name (printed)

 — Signed

05/01/09 — Date

To whom it may concern.

I, the undersigned, give permission for the scholarly use of a manuscript of which I am a co-author in the Doctoral thesis of Jennifer Anne Hutcheon, McGill University Department of Epidemiology, Biostatistics, and Occupational Health:

"Customised birth weight percentiles: does adjusting for maternal characteristics matter?" BJOG 2008; 115:1397–1404.

I understand that this manuscript is intended for inclusion in a thesis by manuscript according to the rules published by McGill University at:
http://www.mcgill.ca/gps/current/programs/thesis/guidelines/preparation/

Under these rules, I waive any issues of copyright related to the inclusion of this manuscript in the Doctoral thesis.

Sincerely,

Michael S. Kramer — Name (printed)

Signed

22/12/08 — Date

To whom it may concern.

I, the undersigned, give permission for the scholarly use of a manuscript of which I am a co-author in the Doctoral thesis of Jennifer Anne Hutcheon, McGill University Department of Epidemiology, Biostatistics, and Occupational Health:

"Customised birth weight percentiles: does adjusting for maternal characteristics matter?" BJOG 2008; 115:1397–1404.

I understand that this manuscript is intended for inclusion in a thesis by manuscript according to the rules published by McGill University at:
http://www.mcgill.ca/gps/current/programs/thesis/guidelines/preparation/

Under these rules, I waive any issues of copyright related to the inclusion of this manuscript in the Doctoral thesis.

Sincerely,

Sven Knutsen — Name (printed)

Signed

December 16, 2008 — Date

To whom it may concern.

I, the undersigned, give permission for the scholarly use of a manuscript of which I am a co-author in the Doctoral thesis of Jennifer Anne Hutcheon, McGill University Department of Epidemiology, Biostatistics, and Occupational Health:

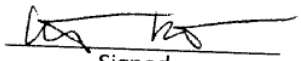
"Customised birth weight percentiles: does adjusting for maternal characteristics matter?" BJOG 2008; 115:1397–1404.

I understand that this manuscript is intended for inclusion in a thesis by manuscript according to the rules published by McGill University at:
http://www.mcgill.ca/gps/current/programs/thesis/guidelines/preparation/

Under these rules, I waive any issues of copyright related to the inclusion of this manuscript in the Doctoral thesis.

Sincerely,

ROBERT PLATT
Name (printed)


Signed

05/01/09
Date

To whom it may concern.

I, the undersigned, give permission for the scholarly use of a manuscript of which I am a co-author in the Doctoral thesis of Jennifer Anne Hutcheon, McGill University Department of Epidemiology, Biostatistics, and Occupational Health:

"Customised birth weight percentiles: does adjusting for maternal characteristics matter?" BJOG 2008; 115:1397–1404.

I understand that this manuscript is intended for inclusion in a thesis by manuscript according to the rules published by McGill University at:
http://www.mcgill.ca/gps/current/programs/thesis/guidelines/preparation/

Under these rules, I waive any issues of copyright related to the inclusion of this manuscript in the Doctoral thesis.

Sincerely,

XUN ZHANG
Name (printed)

Signed

Dec. 16, 2008
Date

Appendix G. Reprints of published manuscripts

1. Hutcheon JA, Platt RW. The missing data problem in birth weight percentiles and thresholds for “small-for-gestational-age”. Am J Epidemiol 2008;167:786-792.

2. Paneth N. Invited Commentary: the hidden population in perinatal epidemiology. Am J Epidemiol 2008;167:793-796.

3. Hutcheon JA, Platt RW. Hutcheon and Platt respond to “the hidden population in perinatal epidemiology”. Am J Epidemiol 2008;167:797-798.

4. Hutcheon JA, Zhang X, Cnattingius S, Kramer MS, Platt RW. Customised birthweight percentiles: does adjusting for maternal characteristics matter?

American Journal of Epidemiology
Vol. 167, No. 7
DOI: 10.1093/aje/kwm327
Advance Access publication March 14, 2008

Practice of Epidemiology

The Missing Data Problem in Birth Weight Percentiles and Thresholds for "Small-for-Gestational-Age"

Jennifer A. Hutcheon[1] and Robert W. Platt[1,2]

[1] Department of Epidemiology and Biostatistics, McGill University Faculty of Medicine, Montreal, Quebec, Canada.
[2] Department of Pediatrics, McGill University Faculty of Medicine, Montreal, Quebec, Canada.

Received for publication August 10, 2007; accepted for publication October 9, 2007.

Weight-for-gestational-age charts and definitions of "small-for-gestational-age" based on the distribution of livebirths at a given gestational age have conventionally been used to identify infants whose fetal growth is poor. However, references based on the weights of only livebirths have serious shortcomings at preterm ages due to missing data on the weights of fetuses still in utero, and these missing data introduce considerable bias to etiologic studies of fetal growth restriction. Application of standard epidemiologic approaches for missing data is needed to help produce perinatal weight percentiles that provide unbiased assessment of fetal growth and risks of small-for-gestational-age.

bias (epidemiology); fetal growth retardation; infant, small for gestational age; reference values

Abbreviations: AGA, appropriate-for-gestational-age; SGA, small-for-gestational-age.

Editor's note: *An invited commentary on this article appears on page 793, and the authors' response is published on page 797.*

Since their first publication by Lubchenco et al. (1) over 40 years ago, weight-for-gestational-age charts (2–5) have been a cornerstone of screening for infants whose intrauterine development is poor. In these charts, the weight distributions of livebirths at each week of gestation are converted to percentiles, and any infant whose weight falls below a certain statistical threshold of the population, typically the 10th percentile, is labeled as being "small-for-gestational-age" (SGA) and is considered to be at increased risk of perinatal morbidity and mortality (6–8). SGA, an anthropometric characteristic that does not necessarily have any adverse health implications, is therefore commonly (but not necessarily appropriately) used as a proxy for the pathologic outcomes believed to be associated with an inadequate rate of fetal growth (9). Weight-for-age charts are nevertheless considered an improvement over low and very low birth weight cutoffs because they differentiate between infants who are small because they are born early in gestation and infants born later but small relative to their peers (10).

In addition to clinical use for the identification of high-risk infants, weight-for-gestational-age charts and their resulting thresholds to define SGA are frequently used in epidemiologic studies of fetal (intrauterine) growth restriction (11–16). Because fetal growth restriction is typically not measurable in population-based data, the majority of research to identify risk factors for fetal growth restriction consists of comparisons of the risks of SGA among exposed and unexposed groups of infants. Although case definitions for SGA or "appropriate-for-gestational-age" (AGA) established by using conventional weight-for-gestational-age charts are well accepted in perinatal epidemiology, their validity according to general epidemiologic principles has rarely been considered.

Correspondence to Jennifer A. Hutcheon, The Montreal Children's Hospital Research Institute, 4060 Rue Sainte Catherine Ouest #205, Westmount, Quebec, Canada H3Z 2Z3 (e-mail: jennifer.hutcheon@mail.mcgill.ca).

The purpose of this article is to argue that size-for-gestational-age charts created from the weight distributions of livebirths have serious shortcomings due to missing data on the weights of fetuses that remain in utero at each gestational age. We will demonstrate how use of a case definition for SGA produced by these charts can introduce considerable bias to estimates of relative risk in etiologic studies of fetal growth restriction, and we will propose that standard statistical and epidemiologic approaches to missing data could be applied to address the bias that currently exists in this field of research.

THE MISSING DATA IN BIRTH WEIGHT REFERENCES

In the study of fetal growth, perinatal researchers have traditionally been faced with what is, in essence, a ''missing data'' problem (17, 18). Although the biological process of interest is the changing fetal size throughout pregnancy, this process is generally unobservable at a population level. In a cohort of conceptions followed forward in time, information on fetal size at any given week of pregnancy is readily available for only the portion of the cohort born during that week. The weights of the remainder of the conception cohort at that gestational age, the fetuses still in utero, are unavailable (19). Although prenatal ultrasonography has enabled estimation of fetal weight prior to the time of birth (20, 21), this information has not been incorporated into reference charts because of concerns over measurement error (2). National weight-for-gestational-age charts (2–5) therefore continue to be created from the weight distributions of only livebirths at each age and are missing the weights of fetuses in the conception cohort not yet born by the end of a given gestational week. In these charts, calculation of weight percentiles at early gestational ages is based on the weights of an extremely small fraction of the total cohort at risk (since the vast majority remain in utero at preterm ages), and, even for gestational ages as old as 36 weeks, weight data for more than 97 percent of the original cohort are still missing (2).

As with any missing-data situation in epidemiology, the extent to which bias will be introduced by the missing intrauterine weights will depend on the mechanism that caused the missingness (22–24). In order for the ''complete case'' approach (25) used in conventional weight-for-gestational-age charts to be valid (the use of only those cases for whom complete data are available—i.e., data on livebirths), the unobserved data must be missing completely at random (MCAR). For data to be missing completely at random, they must represent a randomly selected subset of the total cohort at risk. At term, when the weight distribution of those born at a given gestational age (such as 39 weeks) is likely fairly representative of those still in utero (those who will be born at 40 weeks or later), the assumption of being ''missing-completely-at-random'' may be reasonably valid, and there will be minimal bias from the missing data on intrauterine weights. At preterm ages, however, the available weights are most likely *not* a random sample of the weight distribution of the total at-risk population. Intrauterine growth restriction is a common indication for medically necessary preterm birth

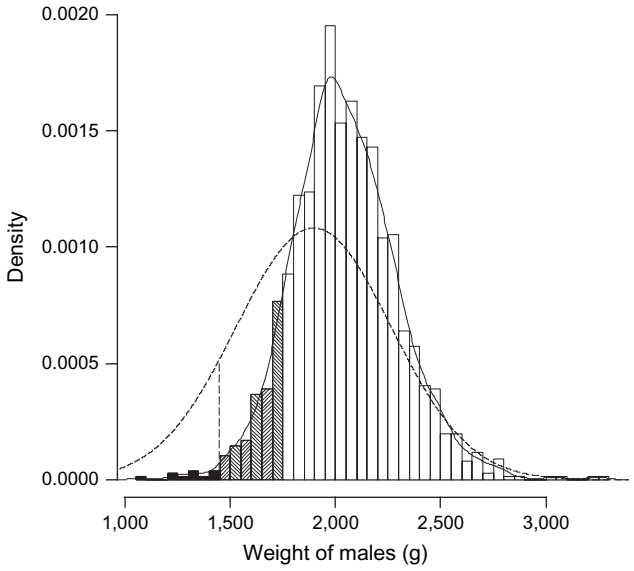


FIGURE 1. Comparison of the estimated fetal weight distribution of male fetuses aged 32 weeks in a general obstetric population at the Royal Victoria Hospital in Montreal, Canada, 2001–2004 (clear bars, $n = 1{,}540$; mean = 2,043 g, standard deviation, 259) with the distribution of a Canadian livebirth weight reference for males aged 32 weeks (curved dashed line) (2). Hatched shading indicates the smallest 10% of the in utero population, black shading indicates those in the smallest 10th percentile of the in utero population identified as small-for-gestational-age by the livebirth reference (threshold for small-for-gestational-age is indicated by the vertical dashed line). Refer to the text for further details.

(26, 27), and the observation that preterm livebirths are smaller than their in utero peers (28–33) has led to speculation that there may be a common cause of spontaneous preterm birth and poor growth. As a result, the observed distribution of weights at earlier gestational ages is systematically shifted to lower values than the weight distribution of the remainder of the cohort at risk at the start of that gestational week. The ''complete-case'' approach used in existing reference charts, when the missing data are clearly not missing completely at random, is therefore inappropriate.

Recognition of the ''missing data'' problem in weight-for-gestational-age charts is certainly not new. Even with the introduction of the first neonatal weight percentile charts in 1963, Lula Lubchenco warned that ''the sample has an undeterminable bias because premature birth itself is probably related to unphysiological states of variable duration in either mother or fetus. Since the weight of fetuses that remain in utero cannot be measured, the curves presented herein are submitted with these reservations . . . '' (1, p. 793). Differences between the weights of preterm livebirths and their in utero peers may be well acknowledged, but what does not appear to have widespread appreciation is the *extent* and *impact* of the bias that the missing data introduce.

The major discrepancy between intrauterine and livebirth weight distributions at preterm ages, as reported in previous publications (30–33), is illustrated in figure 1. In the figure,

the distributions of estimated fetal weights (20) of male singletons aged 32 weeks in an unselected obstetric population at the Royal Victoria Hospital, a McGill University teaching hospital in Montreal, Canada (unpublished data), are compared with a Canadian birth weight reference (2). These ultrasound data are from the years 2001–2004 and were obtained through an institutional policy of universal 32-week ultrasound examinations. The median estimated weight of the fetuses still in utero is more than 120 g heavier than that of livebirths, while the 10th percentile (SGA) threshold of the intrauterine population is more than 300 g higher than the 10th percentile of the national birth weight reference. Similar results were obtained for female fetuses (data not shown). This discrepancy between the 10th percentile thresholds of the two distributions means that applying the national birth weight reference to the intrauterine population (which, at this age, constitutes >99.7 of the total conception cohort (2)) will not identify 10 percent of the population as SGA. Instead, since the 10th percentile of the national birth weight reference is much lower than the 10th percentile of the total cohort, the livebirth weight-based reference will identify less than 1 percent of the total cohort as SGA. That is, the SGA threshold produced by a national birth weight reference at 32 weeks will capture the smallest 1 percent of the total cohort instead of the smallest 10 percent. Error arising from the use of a formula to estimate fetal weight will introduce some bias to estimates of the discrepancy between in- and ex-utero weight distributions; however, since most of this error is random, not systematic (20), it is unlikely to explain a major portion of the discrepancy. The genuine discrepancy between the weights of the in- and ex-utero populations is supported by work such as Hediger et al.'s (28), who demonstrated that the 32-week estimated weights of fetuses that were later born preterm were significantly lower than the 32-week estimated fetal weights of those that were subsequently born at term.

IMPACT OF BIASED WEIGHT PERCENTILES ON PERINATAL EPIDEMIOLOGY

Birth weight references that ignore the missing data of intrauterine weights introduce considerable bias into epidemiologic studies of the etiology of fetal growth restriction. In many studies (11–16), the effect of potential risk factors on fetal growth restriction is evaluated by establishing the relative risk of being SGA between exposed and unexposed infants, calculated as

$$\text{Relative risk}_{\text{SGA}} = (\text{no. SGA}_{\text{exposed}}/\text{no. at-risk}_{\text{exposed}})/(\text{no. SGA}_{\text{unexposed}}/\text{no. at-risk}_{\text{unexposed}}). \quad (1)$$

If the outcome, SGA, is established using a conventional reference based on the distribution of livebirths, an infant's chance of being classified as SGA will change according to his or her gestational age at birth. Consider the case of a male fetus weighing 1,650 g at 32^0 weeks of gestation (32 weeks, zero days). When the weight of this fetus is compared with the weights of all pregnancies that progressed to 32 weeks, this fetus is at approximately the 5th percentile of the population (based on figure 1 data). If all pregnancies in this cohort continued at similar relative growth rates (i.e., the rank order of weights remains unchanged) until birth at 40 weeks' gestation (term), this infant would be classified as SGA by virtue of being in the smallest 10 percent of the population of births at 40 weeks. However, if this fetus were instead born the following day, at 32^1 weeks of gestation, under existing national birth weight references (2) it would instead be assigned the 25th percentile and considered AGA. Although its weight relative to that of its peers of similar gestational stage is constant, classification of this infant as AGA or SGA will be different according to the timing of its birth. A younger gestational age at birth therefore becomes "protective" against being classified as SGA when based on a reference created from weights of livebirths.

The impact of gestational age at birth on the criteria for being defined as SGA becomes problematic in perinatal epidemiology because many risk factors for growth restriction (e.g., smoking, preeclampsia/pregnancy-induced hypertension, multiple births, and disadvantaged ethnicity (34–39)) have also been found to be associated with a younger gestational age at birth or increased rate of preterm birth. As a result, this leads to a differential case definition of SGA being applied to exposed and unexposed groups. Exposed infants are more likely to be born at a younger gestational age, and, at younger gestational ages, the threshold to be identified as SGA is more stringent. At 32 weeks, for example, an infant must be among the smallest 1 percent of his or her remaining conception cohort to be labeled SGA, while, at 40 weeks, the infant need be among only the smallest 10 percent. This difference results in relatively fewer exposed infants being classified as SGA compared the unexposed group, for whom the threshold for SGA is less stringent because of older mean age at birth. As evident from equation 1, an underdiagnosis of $\text{SGA}_{\text{exposed}}$ infants will result in underestimation of the risk of SGA among the exposed and an underestimation or potentially even a reversal of the true measure of effect.

The amount of bias introduced because of differential misclassification of preterm SGA neonates as AGA can be quantified through a simple simulation (table 1). To begin, an estimate of the relative risk of SGA among newborns exposed to preeclampsia (compared with normotensive pregnancies) determined by using a livebirth reference was obtained from previously published research (13), along with the mean gestational ages at birth in each exposure group. The reported unadjusted relative risk of SGA was 2.72, with a mean gestational age among the unexposed of 39.0 weeks (standard deviation, 2.3) and a mean gestational age among the exposed of 37.4 weeks (standard deviation, 3.4). These values were used to generate cohorts of 10,000 exposed and 10,000 unexposed newborns. For each gestational age, the percentage of infants whose weight was in the smallest 10 percent of the total cohort, but not of livebirths, was calculated (i.e., the percentage of SGA infants misclassified as AGA because of the use of a reference based on livebirths was established). The percentage of misclassification at each gestational age was determined by comparing the 10th percentile thresholds of a Norwegian birth weight reference (5) and a Norwegian longitudinal ultrasound

TABLE 1. Bias to the estimate of relative risk of SGA* introduced because of misclassification of preterm SGA infants in a simulated population of 10,000 exposed and 10,000 unexposed infants, with the percentage of misclassification based on Norwegian population data

Gestational age at birth (weeks)	% of SGA infants misclassified as AGA*,†	Unexposed cohort (gestational age = 39.0 weeks (standard deviation, 2.3))‡			Exposed cohort (gestational age = 39.0 weeks (standard deviation, 2.3))‡		
		No. of births	Observed no. of SGA births§	Corrected no. of SGA births	No. of births	Observed no. of SGA births¶	Corrected no. of SGA births
25	65.5				1	0.3	0.8
26	70.2				1	0.3	0.9
27	72.56				11	3.0	10.9
28	71.5				17	4.6	16.2
29	71.0	1	0.1	0.3	66	18.0	61.9
30	65.1	1	0.1	0.3	134	36.4	104.4
31	59.0	7	0.7	1.7	225	61.2	149.3
32	52.0	29	2.9	6.0	358	97.4	202.9
33	43.0	89	8.9	15.6	602	163.7	287.3
34	30.1	274	27.4	39.2	793	215.7	308.6
35	23.0	534	53.4	69.4	1,099	298.9	388.2
36	13.9	999	99.9	116.0	1,246	338.9	393.6
37	0	1,341	134.1	134.1	1,188	323.1	323.1
38	0	1,723	172.3	172.3	1,159	315.2	315.2
39	0	1,688	168.8	168.8	995	270.6	270.6
40	0	1,413	141.3	141.3	776	211.0	211.1
41	0	964	96.4	96.4	558	151.8	151.8
42	0	558	55.8	55.8	350	95.2	95.2
43	0	238	23.8	23.8	211	57.4	57.4
Total		9,859	985.9	1,041.07	9,790	2,662.9	3,349.46
Risk of SGA per 100			10	10.6		27.2	34.2
Relative risk of SGA#						2.72	3.24

* SGA, small-for-gestational-age; AGA, appropriate-for-gestational-age.

† Established by calculating the percentage of infants at preterm ages below the 10th percentile of a Norwegian intrauterine weight reference (42) who were not identified as SGA by a Norwegian birth weight reference (5).

‡ Cohort distribution was truncated at 43 weeks of gestation; therefore, the total number of births does not add to 10,000.

§ Number of SGA births observed with use of a livebirth weight-based reference assuming that the risk of SGA is 10% among the unexposed.

¶ Number of SGA births observed with use of a livebirth weight-based reference if the observed relative risk for the exposed is 2.72.

Compared with that for the unexposed cohort.

reference (40) prior to 37 weeks, at which age the misclassification was zero.

The percentage of misclassifications was then used to ''correct'' the number of SGA cases at each gestational age for both exposed and unexposed groups. As expected, the number of SGA cases increased more in the exposed than in the unexposed group following the correction, since the younger mean gestational age at birth among the exposed would make them more subject to misclassification as AGA. The relative risk of SGA among the exposed was recalculated with the corrected number of SGA cases. The relative risk of SGA of 2.72 presented in the original study when a livebirth reference was used was recalculated to a relative risk of 3.24, a nontrivial difference in effect size that creates a real possibility that true effects of exposures could be found nonsignificant or even potentially reversed because of the differential misclassification of SGA infants. For example, had a relative risk of 0.8 been found with the use of a livebirth reference, the true measure of effect would actually likely be a nearly null effect (relative risk = 0.95 based on Norwegian data, calculations not shown). Covariate adjustment for gestational age as a means to correct this problem is not appropriate, since stratification by gestational age is similar to calculating gestational-age-specific hazards

with a denominator of livebirths, instead of fetuses at risk (41).

The differential misclassification not only will affect observed measures of effect but could also create apparent, but likely spurious, biologic interactions. In a recent study, the relation between SGA birth in a first pregnancy and risk of stillbirth in a subsequent pregnancy was examined (42). The authors reported that the risk of stillbirth in a second pregnancy increased with decreasing gestational age at birth of an SGA infant in the woman's first pregnancy (odds ratio of stillbirth after ''very preterm SGA birth'' > odds ratio after ''preterm SGA birth'' > odds ratio after ''term SGA birth'' when compared with AGA of all ages) and concluded that ''interestingly, the results in this study also reveal that SGA should be considered a heterogenous disease in terms of risk amplitude for subsequent stillbirth. A woman with a term SGA in an index pregnancy is at lower risk level than her counterpart who experiences a preterm SGA, and the greatest risk for stillbirth occurs in women with very preterm SGA'' (42, p. 855). Before concluding that there may be effect modification in the effects of SGA on the risk of stillbirth by gestational age at birth, the potential impact of the bias from livebirth references in this study should be considered. Because those ''very preterm infants'' classified as SGA were in approximately the lowest 1 percent of their conception cohort (based on figure 1 data), whereas the SGA infants at term were in only the lowest 10 percent, it is perhaps not surprising that the more severe cases of growth restriction that consisted of the ''very preterm'' group were found to be a marker for a much greater risk of subsequent stillbirth.

CORRECTING THE MISSING DATA BIAS

To correct the missing data problem in weight-for-gestational-age charts, epidemiologic methods for missing data that are consistent with the nature of the missingness should be applied (22, 23). At preterm ages, the missing data in neonatal weight references are clearly not missing completely at random, making the current ''complete-case'' approach inappropriate. If the distribution of missing intrauterine weights were similar to that of the available birth weight data within strata of known covariates (i.e., if we were able to predict the missingness based on known covariate information), the data would be missing at random (MAR). With missing-at-random data, approaches such as multiple imputation (23) or inverse weighting (43) could be used to build references that accounted for the missing weights. However, since our ability to explain the missingness (amounting to predicting gestational age at birth) is generally agreed to be poor, even considering all known social and medical risk factors (44), these data are likely not missing at random. The missing data in neonatal weight references are therefore likely missing not at random (MNAR), meaning that the missingness process depends on unobserved variables, and any weight-for-gestational-age reference must take this missing data mechanism into account.

A variety of attempts to address the bias from missing intrauterine weights have been proposed in the literature, but none have appropriately addressed the missing-not-at-random nature of the data. Population references based on the distribution of estimated fetal weights (32, 40, 45) represent an improvement at preterm ages (46, 47) but, later in gestation, will introduce missing data bias of their own because of missing weights for those in the population who have been born. ''Hybrid'' references, which either average the growth curves created from livebirth and intrauterine weights (48) or switch from intrauterine weight distributions to birth weight distributions at 37 weeks (49), have also been proposed. While correct in spirit, neither of these approaches accurately reflects the portions of the population in- and ex-utero at each gestational age.

Although options for analyzing missing-not-at-random data are usually limited (22, 23), the case of neonatal weight references represents a relatively rare situation in which external data can be incorporated to produce valid results. With missing-not-at-random data, the weight distributions will be different between those with and without missing data, even within strata of observed covariates. Here, estimates of the weight distributions of those with missing data can be obtained from estimates of fetal weight produced by obstetrical ultrasound (20, 21). Although such estimates have a considerable amount of random error (50), this problem is mainly of concern for predicting weight at the individual level, not for the weight distributions of the population as a whole. Since the magnitude and direction of error in estimates of fetal weight have been reported in validation studies for fetal weight formulae (20, 21), correction for error (both systematic and random) when estimating the weight distribution of the population with missing data should be feasible by using simple Bayesian methods (51). Information on weight distributions of the in- and ex-utero portions of the population can therefore be combined to simulate a cohort with the weights of all fetuses at risk at the beginning of each gestational week.

CONCLUSIONS

The missing intrauterine weight data in conventional birth weight references have resulted in a case definition for SGA that reflects ''a birth weight below the population 10th percentile, corrected for gestational age'' (52, p. 870) only at term ages. At preterm ages, the threshold for SGA reflects a much lower percentage of the total at-risk population, leading to a case definition of SGA that is inconsistent across gestational ages. This case definition is problematic for epidemiologic studies, where exposures of interest are often associated with gestational duration and therefore can affect case status through mechanisms independent of their effect on weight. To correct the missing-data bias that currently exists in studies of fetal growth restriction, the following changes are needed:

1. References to assess neonatal weight must be developed that reflect the weight distributions of all fetuses in the population at the beginning of a given gestational week, not just livebirths. By definition, preterm births are not ''normal'' pregnancies and should therefore not be used to characterize the growth patterns of the full conception

cohort. Thus, the correct reference chart is neither a livebirth weight reference nor an intrauterine estimated fetal weight reference, but a *perinatal* one that combines the weights of both livebirths *and* fetuses in utero at each week of gestation. At preterm ages, it will constitute predominantly in utero weights; as term ages approach, livebirth weights will make up a larger and larger portion of the distribution.

2. The case definition of SGA must be established as the bottom 10 percent of the total population at risk of being small, not just those who happen to be born at a given week of gestation. To establish that an infant of 32 weeks is small for its gestational age, its size needs to be compared with that of all other pregnancies that progressed to 32 weeks, regardless of whether those pregnancies went on to end at 32 weeks or 40 weeks. Researchers should stop classifying as normal the weights of growth-restricted preterm infants simply because there are many other growth-restricted preterm livebirths who are even smaller than they are. This case definition is particularly important for etiologic studies of growth restriction, to prevent differential misclassification of SGA cases as noncases. Until an unbiased reference is available, the use of birth-weight-for-gestational-age charts should be restricted to term ages (53).

Adopting the same approaches to the missing data in neonatal weight charts as we would for missing data in other areas of epidemiology will likely do much to further our understanding of perinatal population health.

ACKNOWLEDGMENTS

J. A. H. is supported by a Canadian Institutes of Health Research (CIHR) Doctoral Research Award, and R. W. P. holds a Chercheur-Boursier from the Fonds de la Recherche en Santé Quebec (FRSQ).

The authors thank Dr. Maala Bhatt for comments on earlier versions of the manuscript and Dr. Jean-François Boivin for helpful suggestions following a seminar presentation of this work.

Conflict of interest: none declared.

REFERENCES

1. Lubchenco LO, Hansman C, Dressler M, et al. Intrauterine growth as estimated from liveborn birth-weight data at 24 to 72 weeks of gestation. Pediatrics 1963;32:793–800.
2. Kramer MS, Platt RW, Wen SW, et al. A new and improved population-based Canadian reference for birth weight for gestational age. (Electronic article). Pediatrics 2001; 108:e35.
3. Källén B. A birth weight for gestational age standard based on data in the Swedish Medical Birth Registry, 1985–1989. Eur J Epidemiol 1995;11:601–6.
4. Alexander G, Himes J, Kaufman R, et al. A United States national reference for fetal growth. Obstet Gynecol 1996;87: 163–8.
5. Skjaerven R, Gjessing HK, Bakketeig LS. Birthweight by gestational age in Norway. Acta Obstet Gynecol Scand 2000;79:440–9.
6. Battaglia FC, Lubchenco LO. A practical classification of newborn infants by weight and gestational age. J Pediatr 1967;71:159–63.
7. Koops BL, Battaglia FC. Neonatal mortality risk in relation to birth weight and gestational age: update. J Pediatr 1982;101: 969–77.
8. Wilcox AJ, Skjaerven RS. Birth weight and perinatal mortality: the effect of gestational age. Am J Public Health 1992;82: 378–82.
9. Altman DG, Hytten FE. Intrauterine growth retardation: let's be clear about it. Br J Obstet Gynaecol 1989;96:1127–32.
10. Arnold CC, Kramer MS, Hobbs CA, et al. Very low birth weight: a problematic cohort for epidemiologic studies of very small or immature neonates. Am J Epidemiol 1991;134: 604–13.
11. Kornfeld D, Cnattingius S, Ekbom A. Pregnancy outcomes in women with inflammatory bowel disease—a population-based cohort. Am J Obstet Gynecol 1997;177:942–6.
12. Ahluwalia IB, Grummer-Strawn L, Scanlon KS. Exposure to environmental tobacco smoke and birth outcome: increased effects on pregnant women aged 30 years or older. Am J Epidemiol 1997;146:42–7.
13. Xiong X, Demianczuk NN, Buekens P, et al. Association of preeclampsia with high birth weight for gestational age. Am J Obstet Gynecol 2000;183:148–55.
14. Liu S, Wen SW, Demissie K, et al. Maternal asthma and pregnancy outcomes: a retrospective cohort study. Am J Obstet Gynecol 2001;184:90–6.
15. Buchbinder A, Sibai BM, Caritis S, et al. Adverse pregnancy outcomes are significantly higher in severe gestational hypertension than in mild pre-eclampsia. Am J Obstet Gynecol 2002;186:66–71.
16. Luo ZC, Simonet F, An N, et al. Effect on neonatal outcomes in gestational hypertension in twin compared with singleton pregnancies. Obstet Gynecol 2006;108:1138–44.
17. Platt RW, Joseph KS, Ananth CV, et al. A proportional hazards model with time-dependent covariates and time-varying effects for analysis of fetal and infant death. Am J Epidemiol 2004;160:199–206.
18. Hernán MA, Schisterman EF, Hernández-Díaz S, et al. Prenatal exposures, gestational age, birth weight, and mortality: an application of causal diagrams. (Abstract). In: Proceedings of the 17th Annual Meeting of the Society for Pediatric and Perinatal Epidemiologic Research, Salt Lake City, Utah, June 2004. Ann Arbor, MI: Society for Pediatric and Perinatal Epidemiologic Research, 2004:38.
19. Savitz DA, Hertz-Picciotto I, Poole C, et al. Epidemiologic measures of the course and outcome of pregnancy. Epidemiol Rev 2002;24:91–101.
20. Hadlock FP, Harrist RB, Sharman RS, et al. Estimation of fetal weight with the use of head, body, and femur measurement—a prospective study. Am J Obstet Gynecol 1985;151:333–7.
21. Shepard MJ, Richards VA, Berkowitz RL, et al. An evaluation of two equations for predicting fetal weight by ultrasound. Am J Obstet Gynecol 1982;142:47–54.
22. Raghunathan TE. What do we do with missing data? Some options for analysis of incomplete data. Annu Rev Public Health 2004;25:99–117.
23. Schafer JL, Graham JW. Missing data: our view of the state of the art. Psychol Methods 2002;7:147–77.
24. Rubin DB. Inference and missing data. Biometrika 1976;63: 581–92.

25. Greenland S, Finkle WD. A critical look at methods for handling missing covariates in epidemiologic regression analyses. Am J Epidemiol 1995;142:1255–64.
26. Papiernik E. Fetal growth retardation: a limit for the further reduction of preterm births. Matern Child Health J 1999;3:63–9.
27. Joseph KS, Demissie K, Kramer MS. Obstetric intervention, stillbirth, and preterm birth. Semin Perinatol 2002;26:250–9.
28. Hediger ML, Scholl TO, Schall JI, et al. Fetal growth and the etiology of preterm delivery. Obstet Gynecol 1995;85:175–82.
29. Weiner CP, Sabbagha RE, Vaisrub N, et al. A hypothetical model suggesting suboptimal intrauterine growth in infants delivered preterm. Obstet Gynecol 1985;65:323–5.
30. Secher NJ, Hansen K, Thomsen BL, et al. Growth retardation in preterm infants. Br J Obstet Gynaecol 1987;66:463–71.
31. Ott WJ. Intrauterine growth retardation and preterm delivery. Am J Obstet Gynecol 1993;168:1710–15.
32. Marsal K, Persson PH, Larsen T, et al. Intrauterine growth curves based on ultrasonically estimated foetal weights. Acta Paediatr 1996;85:843–8.
33. Fry AG, Bernstein IM, Badger GJ. Comparison of fetal growth estimates based on birth weight and ultrasound references. J Matern Fetal Neonatal Med 2002;12:247–52.
34. Wilcox AJ. Birth weight and perinatal mortality: the effect of maternal smoking. Am J Epidemiol 1993;137:1098–104.
35. Shiono PH, Klebanoff MA, Rhoads GG. Smoking and drinking during pregnancy: their effect on preterm birth. JAMA 1986;255:82–4.
36. Joseph KS, Kramer MS, Marcoux S, et al. Determinants of preterm birth rates in Canada from 1981 through 1983 and 1992 through 1994. N Engl J Med 1998;339:1434–9.
37. Min SJ, Luke B, Gillespie B, et al. Birth weight references for twins. Am J Obstet Gynecol 2000;182:1250–7.
38. Demissie K, Rhoads GG, Ananth CV, et al. Trends in preterm birth and neonatal mortality among blacks and whites in the United States from 1989 to 1997. Am J Epidemiol 2001;154:307–15.
39. Odegard RA, Vatten LJ, Nilsen ST, et al. Pre-eclampsia and fetal growth. Obstet Gynecol 2000;96:950–5.
40. Johnsen SL, Rasmussen S, Wilsgaard T, et al. Longitudinal reference ranges for estimated fetal weight. Acta Obstet Gynecol Scand 2006;85:286–97.
41. Joseph KS. Incidence-based measures of birth, growth restriction, and death can free perinatal epidemiology from erroneous concepts of risk. J Clin Epidemiol 2004;57:889–97.
42. Salihu HM, Sharma PP, Aliyu MH, et al. Is small for gestational age a marker of future fetal survival in utero? Obstet Gynecol 2006;107:851–6.
43. Robins JM, Rotnitzky A, Zhao LP. Estimation of regression coefficients when some regressors are not always observed. J Am Stat Assoc 1994;89:846–66.
44. Buekens P, Klebanoff M. Preterm birth research: from disillusion to the search for new mechanisms. Paediatr Perinat Epidemiol 2001;15:159–61.
45. Hadlock FP, Harrist RB, Martinez-Poyer J. In utero analysis of fetal growth: a sonographic weight standard. Radiology 1991;181:129–33.
46. Zaw W, Gagnon R, da Silva O. The risks of adverse neonatal outcome among preterm small for gestational age infants according to neonatal versus fetal growth standards. Pediatrics 2003;111:1273–7.
47. Lackman F, Capewell V, Richardson B, et al. The risks of spontaneous preterm delivery and perinatal mortality in relation to size at birth according to fetal versus neonatal growth standards. Am J Obstet Gynecol 2001;184:946–53.
48. Mongelli M, Biswas A. A fetal growth standard derived from multiple modalities. Early Hum Dev 2001;60:171–7.
49. Bernstein IM, Mohs G, Rucquoi M, et al. Case for hybrid ''fetal growth curves'': a population-based estimation of normal fetal size across gestational age. J Matern Fetal Med 1996;5:124–7.
50. Dudley NJ. A systematic review of the ultrasound estimation of fetal weight. Ultrasound Obstet Gynecol 2005;25:80–9.
51. Gelman A, Carlin JB, Stern HS, et al. Bayesian data analysis. 2nd ed. London, United Kingdom: Chapman & Hall/CRC, 2003.
52. Janssen PA, Thiessen P, Klein MC, et al. Standards for the measurement of birth weight, length, and head circumference at term in neonates of European, Chinese, and South Asian ancestry. Open Med 2007;1:74–88.
53. Gabbe SG, Niebyl JR, Leigh Simpson J, eds. Obstetrics: normal and problem pregnancies. 4th ed. Orlando, FL: Churchill Livingstone Inc, 2002.

American Journal of Epidemiology

Vol. 167, No. 7
DOI: 10.1093/aje/kwn041
Advance Access publication March 14, 2008

Invited Commentary

Invited Commentary: The Hidden Population in Perinatal Epidemiology

Nigel Paneth[1,2]

[1] Department of Epidemiology, College of Human Medicine, Michigan State University, East Lansing, MI.
[2] Department of Pediatrics and Human Development, College of Human Medicine, Michigan State University, East Lansing, MI.

Received for publication December 5, 2007; accepted for publication December 21, 2007.

Several recent papers have argued that understanding of pathologic perinatal processes may be advanced by considering, in varying ways, the population of fetuses still in utero. Initially invoked by Yudkin et al. (Lancet 1987;1:1192–4) as the optimum denominator for intrapartum stillbirths, fetuses in utero (or "fetuses at risk") have also been of interest because of their patterns of growth, especially in comparison to infants born after varying durations of gestation. The paper by Hutcheon and Platt (Am J Epidemiol 2008;167:786–792) extends work that compares growth in infants born prematurely with that in infants continuing in utero and investigates the biases in analyses that can emerge from failure to consider the selection for impaired fetal growth characteristic of many prematurely born infants. Although the conceptual basis of this perspective is sound, in-utero fetal growth standards from serial ultrasonographic measurements in pregnancy are often based on small and highly selected samples. Some authors have proposed "fetuses at risk" as the appropriate denominator for postnatal phenomena related to premature birth, such as neonatal mortality and cerebral palsy. This application is problematic; in such situations, the denominator population differs from infants with the outcome in not having experienced adjustment to postnatal life, a potentially important determinant of outcome, especially in premature infants. The fetuses-at-risk concept is important in perinatal epidemiology and has implications for obstetric practice, but it must be handled with caution.

bias (epidemiology); fetal growth retardation; gestational age; infant, small for gestational age; pre-eclampsia; pregnancy; premature birth; ultrasonography, prenatal

THE VEIL OF PREGNANCY

The central difficulty of perinatal epidemiology is that so much is hidden from view by the veil of pregnancy. Of the embryonic population that is formed at conception, as many as a third may be lost within a few weeks, often without the mother's awareness (1). The processes that govern these fetal losses, or even the later fetal losses that occur when pregnancy is under medical surveillance, are poorly understood (2). Nor can we directly observe the way in which the fetus grows and develops; even so simple a measure as fetal weight can be inferred only indirectly from other measures. Many of our inferences in perinatal epidemiology are thus derived from what is observable after pregnancy is over. At that point, we can enumerate the number of livebirths and fetal deaths, and we can ascertain deaths occurring in the first month of life. From these three populations, and the relationships among them, we construct the perinatal mortality statistics that we use to monitor maternal and child public health.

This end-of-pregnancy approach has been widely used to assess the mortal pregnancy outcomes at any gestational age. The neonatal mortality rate for babies born between 32 and 33 weeks of gestation is generally derived from all livebirths in that interval and the number of newborn deaths occurring among them. Similarly, stillborn infants born in that gestational week are conventionally denominatored to all births occurring in that week (livebirths plus stillbirths), generating a "stillbirth rate," which, as most "rates" in epidemiology, is not a measure of change over time, but a proportion. However, pregnancies ending prematurely can also be viewed as part of a cohort of births most of whose members are still in utero. From this perspective, it makes sense to consider the population of fetuses not yet

Correspondence to Dr. Nigel Paneth, B 636 West Fee Hall, Michigan State University, East Lansing, MI 48824 (e-mail: paneth@msu.edu).

delivered at any given gestational week as the denominator population at risk of stillbirth in that week. This compelling argument was first put forth by Yudkin et al. (3) in 1987. These authors appear to have been the first to suggest that in addition to livebirths, fetal deaths, and neonatal deaths, a fourth population—fetuses still in utero—needs to be considered in at least some perinatal analyses. They used the term ''stillbirth risk'' as contrasted to ''stillbirth rate'' to describe their approach.

An interesting finding emerged from Yudkin et al.'s (3) approach. Stillbirth *rates* decrease with gestational age, but stillbirth *risks* increase with gestational age. This observation alerts clinicians to anticipate that the risk of stillbirth, especially unexplained stillbirth, rises as term approaches. The appeal of this way of thinking about risk of stillbirth is strong, since stillbirths by definition do not arise from the population of livebirths. As we shall see, however, the concept of fetuses at risk is more applicable to stillbirths than it is to some other perinatal outcomes.

FETAL GROWTH

Hutcheon and Platt (4), in this issue of the *Journal*, take a parallel stance in relation to fetal growth. The hidden population of fetuses still in utero is also viewed as a reference population of interest, but the situation here is more complicated, requiring not just enumeration but measurement. Serial ultrasonographic measurements of fetuses in utero have shown what many clinicians have long suspected, that premature births are selected for poor fetal growth from among the population of fetuses. Moreover, as Hutcheon and Platt show, this selection bias increases with decreasing gestation, making fetal growth restriction and gestational age inversely correlated with each other. They thus have the potential to confound each other in analyses of perinatal health outcomes. Their paper empirically demonstrates this bias, showing how pregnancy phenomena that are inversely correlated with gestational age at birth—preeclampsia, for example—have stronger associations with fetal growth restriction when that growth is assessed on the basis of standards derived from fetuses in utero rather than on standards derived from livebirths, because the latter group is biased toward fetal growth restriction when gestational duration is curtailed.

The authors (4) show, although perhaps do not emphasize quite enough, that, in their data, selection for premature birth is unevenly distributed across the spectrum of fetal weight. Thus, while the median fetal/birth weight at 32 weeks differs between the two weight standards by 120 g, the difference between the 10th percentiles is 300 g, a value that would be even larger were it expressed as standard deviations from the means being compared. Indeed, if we compare the two weight distributions plotted by the authors (their figure 1), we see that it is not the entire weight distribution that is shifted to the left in the liveborn compared with the fetal sample; rather, those fetuses destined to be born early are selectively removed from the lower 75 percent of the distribution. Above that percentile, the two distributions seem nearly identical. It is good to be reminded that we cannot predict the shape of two distributions from their mean or median difference.

Whereas Yudkin et al. (3) used fetuses as the sole denominator for stillbirth risk, Hutcheon and Platt (4) conclude that a hybrid growth standard, based on both fetuses and livebirths, would best be used to determine infants who suffer from ''intrauterine growth restriction.'' Doing so would greatly expand the number of babies defined as experiencing intrauterine growth restriction, from the conventional lightest 5 percent or 10 percent of the distribution, to a considerably higher figure that increases with decreasing gestational age, reaching as high as 70 percent in the very earliest gestations at which infant viability occurs.

DIFFICULTIES TO CONSIDER

The technique of prenatal ultrasonography has proven reasonably good at estimating birth weight from in-utero measurements, but it is not without its flaws. For one thing, gestational age is itself often estimated from the same source, at times creating a circularity in the logic of fetal-weight-for-gestational-age estimation (5, 6). Second, fetal weight is estimated from formulas based on one-dimensional measures such as femoral length or biparietal diameter, whose measurement errors must be magnified when extrapolated to a measure based on three dimensions, such as weight. Third, because repeated ultrasonographic measurements in pregnancy are not simple to arrange, most longitudinal fetal growth standards are based on small and selected samples. The fetal growth standard used by Hutcheon and Platt (4) is from Norway and is based on 634 pregnancies, each scanned three times on average (7). The births are also Norwegian, but they cover a much longer period of time, during which birth weight for gestational age changed, probably as the result of more liberal use of cesarean section (8). More problematic are comparison studies in which, for example, the births are British and the fetuses Swedish (9) or the births are Canadian and the fetuses Texan (10). In these latter two examples, the sample sizes used to establish the fetal growth standards were, respectively, 86 (11) and 392 (12). One has to consider the possibility that the disparate sources of such comparisons might bear some relation to their divergent findings.

Hutcheon and Platt (4), like most authors in this area, echo the clinical focus on one extreme tail of the distribution—weight below the 5th or the 10th percentile for gestational age. However, there is little to suggest that the effects of impaired growth suddenly become manifest below a fixed threshold. More likely, the entire range of growth is of interest, especially when dealing with infants already deemed to be at risk because of preterm birth. I suspect that both clinicians and scientists interested in the preterm infant would gain more useful information from describing each infant's status in relation to the entire spectrum of growth, whether based on fetal, neonatal, or hybrid standards. The full spectrum of fetal growth can be represented as a ratio of the infant's weight to the median for his or her gestational week (termed the fetal growth ratio) (13) or as standard deviations units (z scores) away from the mean weight for the gestational week (14).

BEYOND FETAL LIFE

The two illustrations of the fetuses-at-risk approach discussed thus far—stillbirths and fetal growth—are both intrauterine phenomena determined before birth. However, the suggestion has been made to use this approach for phenomena occurring after birth, such as neonatal death, and even phenomena of perinatal origin diagnosed much later, such as cerebral palsy (15). The difficulty here is that neonatal death and cerebral palsy occur in liveborn infants only. This is not just a technical matter but a recognition that certain experiences are key predisposers to risk, and, absent those experiences, any comparison of risk may be biased.

The traditional assumption, implicit in gestational growth curves based on livebirths, is that premature birth is a random event occurring in an otherwise normal pregnancy that ended too early. We now know that assumption is untenable. On the other hand, the ''fetus-at-risk'' approach carries the assumption that neonatal death or disability would have occurred with equal probability, and at the same point in time, whether the infant was born early or had stayed in utero. That assumption is likewise untenable. The truth must lie somewhere between these two extremes. For neonatal death and disability linked to preterm birth, the manifold problems associated with successful transition to postnatal life require us to place our marker much closer to the traditional assumption that delivery and birth dramatically alter risk. The imperfect ability of postnatal medical care to mimic the environment of the womb in severely preterm infants means that, even with the best of care, the death and disability rate for infants born at 28 weeks of gestation is orders of magnitude higher than for those remaining in utero. This risk difference cannot entirely be due to birth selection for preexisting damage.

Consider a parallel. Livebirths arise from the denominator population of fetuses, but fetuses themselves arise from a denominator population of women of childbearing age. The fetuses-at risk concept might thus be enlarged to include women at risk as the denominator population. Would we be interested in the rate of preeclampsia in women of childbearing age, ignoring the fact of pregnancy? Obviously not. Just as pregnancy is a requirement for risk of preeclampsia, so is livebirth a requirement for risk of neonatal death and cerebral palsy.

WHY ALL THIS MATTERS

The limitations of prenatal growth standards noted above have deterred some neonatologists from using them to replace birth standards (16), but, in any case, clinicians who care for premature infants will, for the purpose of defining a high-risk group, want to compare growth among their liveborn charges. The growth of these infants compared with their more fortunate peers remaining in utero is not really relevant. Nonetheless, clinicians should recognize that the more premature a baby, the more likely he or she is to deviate in growth from fetuses in utero of the same gestational age.

Obstetricians almost certainly incorporate the fetuses-at-risk model into their decision making, whether consciously or not. The concern that further time in the womb might produce a stillbirth or a severely compromised infant implies a recognition that certain risks increase with increasing gestation, the hallmark of the fetuses-at-risk calculation. It is this concern that motivates the principal obstetric intervention, the decision to hasten birth by induction or operative delivery. A more formal model of this thinking has been provided by Joseph (17).

For epidemiologists, recognition and acknowledgment of the hidden perinatal population—the fetuses quietly biding their time in utero and not coming to our attention as vital events—can at times lead to a more precise understanding of the determinants of perinatal health.

ACKNOWLEDGMENTS

The helpful comments of Richard Ehrenkrantz and Alan Leviton are greatly appreciated.

Conflict of interest: none declared.

REFERENCES

1. Wilcox AJ, Weinberg CR, O'Connor JF, et al. Incidence of early loss of pregnancy. N Engl J Med 1988;319:189–94.
2. Goldenberg RL, Kirby R, Culhane JF. Stillbirth: a review. J Matern Fetal Neonatal Med 2004;16:79–94.
3. Yudkin PL, Wood L, Redman CW. Risk of unexplained stillbirth at different gestational ages. Lancet 1987;1: 1192–4.
4. Hutcheon JA, Platt RW. The missing data problem in birth weight percentiles and thresholds for ''small-for-gestational-age.'' Am J Epidemiol 2008;167:786–92.
5. Henricksen TB, Wilcox AJ, Hedegaard M, et al. Bias in studies of preterm and postterm delivery due to ultrasound assessment of gestational age. Epidemiology 1995;6: 533–7.
6. Dietz PM, England LJ, Callaghan WM, et al. A comparison of LMP-based and ultrasound-based estimates of gestational age using linked California livebirth and prenatal screening records. Paediatr Perinat Epidemiol 2007;21(suppl 2): 62–71.
7. Johnsen SL, Rasmussen S, Wilsgaard T, et al. Longitudinal reference ranges for estimated fetal weight. Acta Obstet Gynecol Scand 2006;85:286–97.
8. Skjaerven R, Gjessing HK, Bakketeig LS. Birthweight by gestational age in Norway. Acta Obstet Gynecol Scand 2000;79:440–9.
9. Cooke RWI. Conventional birth weight standards obscure fetal growth restriction in preterm infants. Arch Dis Child Fetal Neonatal Ed 2007;92:189–92.
10. Lackman F, Capewell V, Richardson B, et al. The risks of spontaneous preterm delivery and perinatal mortality in relation to size at birth according to fetal versus neonatal growth standards. Am J Obstet Gynecol 2001; 184:946–53.
11. Marsál K, Persson PH, Larsen T, et al. Intrauterine growth curves based on ultrasonically estimated foetal weights. Acta Paediatr 1996;85:843–8.

12. Hadlock FP, Harrist RB, Martinez-Poyer J. In utero analysis of fetal growth: a sonographic weight standard. Radiology 1991;181:129–33.
13. Kramer MS, Olivier M, McLean FH, et al. Determinants of fetal growth and body proportionality. Pediatrics 1990; 86:18–26.
14. Zhang J, Bowes WA Jr. Birth-weight-for-gestational-age patterns by race, sex, and parity in the United States population. Obstet Gynecol 1995;86:200–8.
15. Joseph KS, Allen AC, Lutfi S, et al. Does the risk of cerebral palsy increase or decrease with increasing gestational age? BMC Pregnancy Childbirth 2003;3:8.
16. Ehrenkranz R. Estimated fetal weights versus birth weights; should the reference intrauterine growth curves based on birth weights be retired? Arch Dis Child Fetal Neonatal Ed 2007;92:161–2.
17. Joseph KS. Theory of obstetrics: an epidemiologic framework for justifying medically indicated early delivery. BMC Pregnancy Childbirth 2007;7:4.

American Journal of Epidemiology
Vol. 167, No. 7
DOI: 10.1093/aje/kwn038
Advance Access publication March 14, 2008

Response to Invited Commentary

Hutcheon and Platt Respond to ''The Hidden Population in Perinatal Epidemiology''

Jennifer A. Hutcheon[1] and Robert W. Platt[1,2]

[1] Department of Epidemiology and Biostatistics, McGill University Faculty of Medicine, Montreal, Quebec, Canada.
[2] Department of Pediatrics, McGill University Faculty of Medicine, Montreal, Quebec, Canada.

Received for publication January 28, 2008; accepted for publication February 4, 2008.

We thank Dr. Paneth for his insightful comments (1) and, in particular, the way in which he has located our work on fetal growth (2) in the context of the larger body of work on fetuses at risk. We would like to respond to some of the concerns raised regarding the measurement of intrauterine growth, as well as touch briefly on the larger issue of selecting denominators in perinatal epidemiology.

We agree entirely with Dr. Paneth's criticism (1) of the focus on dichotomous measures of fetal growth (''small-for-gestational-age'' vs. ''appropriate-for-gestational-age'') instead of continuous ones such as birth weight z scores. However, as with percentiles, we argue that it is important to ensure that the mean weights (and standard deviations) used to calculate z scores are based on the average weights of all fetuses that progressed to a given gestational age, not the average weights of only fetuses subsequently born that week. We agree that epidemiologists would be well served to respect the natural continuum of fetal growth in their attempts to better understand the etiology of growth restriction, but they should do so with a measure that is not associated with gestational age at birth.

The limitations of existing intrauterine weight standards, including small sample sizes and errors in estimation of fetal weight, are important to recognize. We should clarify that the goal of our calculations was not to establish *the* precise extent of the bias arising from the use of conventional birth weight percentiles but rather to help understand whether the magnitude of the theoretical bias we outlined was of substantive importance. We had also repeated our simulations using a Canadian birth weight reference (3) and ultrasound data from the Royal Victoria Hospital population presented in figure 1 (2) (a sample of 3,015 routine 32-week ultrasounds once female fetuses were included, which, of note, is a larger sample than the 1,792 births at 32 weeks used to create the Canadian reference). This method produced a similar, if slightly larger, bias than that obtained from our simulations based on Norwegian data. This finding, combined with the observation that the 32-week estimated fetal weights of fetuses later born preterm in our population were already significantly lower than the estimated fetal weights of fetuses subsequently born at term (confirming previous reports from a smaller study (4)), led us to believe that the bias presented in our paper was unlikely to be solely an artifact from errors in estimation of fetal weight or fetal weight standards.

The exact magnitude of the bias requires further refinement, however, and research to improve estimated fetal weights formulae and create a new fetal weight standard (5) will be a valuable contribution toward this end. Research is also needed to develop a standard that combines the weights of births and ongoing pregnancies (i.e., all fetuses at risk at the beginning of a given week) so that it is methodologically appropriate for all gestational ages.

Since fetal growth (and growth restriction) occurs in utero, we have argued (2) that all fetuses at risk of being small at a given gestational age should be included in the creation of normative weight ranges. Application of this fetuses-at-risk principle to other areas of perinatal epidemiology is more controversial, and we can appreciate arguments against the use of fetuses at risk for outcomes such as cerebral palsy or infant mortality (1). We would propose, however, that much of the debate on fetuses-at-risk versus total-birth reference groups has suffered from attempts to make generic recommendations to cover all research questions. Perhaps it is time to begin considering each situation individually and to establish the most

Correspondence to Jennifer A. Hutcheon, The Montreal Children's Hospital Research Institute, 4060 Rue Sainte Catherine Ouest #205, Westmount, Quebec, Canada, H3Z 2Z3 (e-mail: jennifer.hutcheon@mail.mcgill.ca).

appropriate denominators based on the research question, the outcome of interest, and the hypothesized timing and mechanism of the exposure. As for neonatologists, although their clinical focus is on the care of their liveborn charges and not fetuses in utero, we would still propose that the most appropriate standard to use is that which is best able to identify infants at increased risk of adverse neonatal outcomes, irrespective of whether this is obtained through an intrauterine or birth weight standard. Some work has been done to compare the predictive ability of intrauterine and neonatal weight standards for risk of adverse outcomes such as respiratory morbidity (6, 7), but studies with sufficient statistical precision to conclusively distinguish between the two standards are needed before recommendations for clinical use can be made.

ACKNOWLEDGMENTS

Conflict of interest: none declared.

REFERENCES

1. Paneth N. Invited commentary: The hidden population in perinatal epidemiology. Am J Epidemiol 2008;167:793–6.
2. Hutcheon JA, Platt RW. The missing data problem in birth weight percentiles and thresholds for ''small-for-gestational-age.'' Am J Epidemiol 2008;167:786–92.
3. Kramer MS, Platt RW, Wen SW, et al. A new and improved population-based Canadian reference for birth weight for gestational age. (Electronic article). Pediatrics 2001;108:E35.
4. Hediger ML, Scholl TO, Schall JI, et al. Fetal growth and the etiology of preterm delivery. Obstet Gynecol 1995;85:175–82.
5. National Institutes of Health. US National Institute of Child Health and Human Development. The national standard for normal fetal growth. (http://www.nichd.nih.gov/about/org/despr/studies/preg/normfetalgrowth.cfm). (Accessed January 26, 2008).
6. Zaw W, Gagnon R, Da Silva O. The risks of adverse neonatal outcome among preterm small for gestational age infants according to neonatal versus fetal growth standards. Pediatrics 2003;111:1273–7.
7. Cooke RWI. Conventional birth weight standards obscure fetal growth restriction in preterm infants. Arch Dis Child Fetal Neonatal Ed 2007;92:189–92.

DOI: 10.1111/j.1471-0528.2008.01870.x
www.blackwellpublishing.com/bjog

Customised birthweight percentiles: does adjusting for maternal characteristics matter?

JA Hutcheon,[a] X Zhang,[b] S Cnattingius,[c] MS Kramer,[a,b] RW Platt[a,b]

[a] Department of Epidemiology and Biostatistics and [b] Department of Pediatrics, McGill University, Montréal, Quebec, Canada
[c] Department of Medical Epidemiology, Karolinska Institute, Stockholm, Sweden
Correspondence: Ms JA Hutcheon, The Montreal Children's Hospital Research Institute, 4060 Rue Sainte Catherine Ouest 205, Westmount, Quebec, Canada, H3Z 2Z3. Email jennifer.hutcheon@mail.mcgill.ca

Accepted 29 June 2008.

Objective The objective of this study was to determine whether the improved prediction of risk for perinatal mortality obtained with the use of a customised birthweight standard can also be obtained with the use of a non-customised but intrauterine-based standard.

Design Population-based cohort study.

Setting Sweden.

Population Births in the Swedish Medical Birth Register between 1992 and 2001 (n = 782 303) with complete data on birthweight, gestational age, sex, maternal age, pre-pregnancy body mass index, height, parity, and ethnicity.

Methods We calculated the relative risks (RRs) of stillbirth and early neonatal mortality among small-for-gestational-age (SGA) births as established by (1) a customised standard, (2) a population standard based on birthweights, and (3) a population standard based on a best estimate of intrauterine weights.

Main outcome measures Stillbirth and early neonatal mortality (<7 days).

Results The RRs of stillbirth and early neonatal mortality among SGA births as classified by the intrauterine standard were similar to those among SGA births as classified by the customised standard and much higher than those among SGA births as classified by the birthweight standard.

Conclusions A non-customised but intrauterine-based standard has a similar ability to predict risk for stillbirth and early neonatal mortality as a customised birthweight standard. The process of customising population weight-for-gestational-age standards to account for maternal characteristics does little to improve prediction of perinatal mortality.

Keywords Infant mortality, intrauterine growth retardation, reference standards, small for gestational age, stillbirth.

Please cite this paper as: Hutcheon J, Zhang X, Cnattingius S, Kramer M, Platt R. Customised birthweight percentiles: does adjusting for maternal characteristics matter? BJOG 2008;115:1397–1404.

Introduction

Customised birthweight percentiles, first proposed by Gardosi *et al.*,[1] are weight-for-gestational-age charts that have been individualised to account for maternal influences on fetal growth. By incorporating information generally believed to have a physiological influence on fetal growth (such as maternal height, pre-pregnancy body mass index [BMI], parity, ethnicity, and fetal sex), customised percentiles were designed to better differentiate between infants who are small because their *in utero* growth has been restricted and infants who are small but have reached their individual growth potential.[2] Customised birthweight percentiles have consistently been shown to be superior to conventional birthweight-for-gestational-age percentiles in predicting perinatal morbidity and mortality[3–5] and, as a result, have been recommended for clinical use by practice guidelines of the British Royal College of Obstetricians and Gynaecologists.[6]

Before customised birthweight charts replace conventional weight-for-age charts in clinical practice, however, a better understanding of their properties is needed. Although the reported benefits of customised birthweight percentiles are generally attributed to their adjustment for maternal characteristics, an alternative explanation for their improved ability to predict perinatal morbidity and mortality is possible. Customised birthweight percentiles have a second methodological difference from conventional birthweight-for-gestational-age charts, in addition to adjusting for maternal characteristics. At earlier gestational ages, the customised percentiles are based on Hadlock's proportionality formula,[2] a formula in

which an infant's predicted 'optimal' birthweight is expressed as a proportion of its 'optimal' weight predicted for 280 days (40 weeks) according to the intrauterine growth curve of Hadlock.[7] In essence, the normative values in customised percentiles at younger gestational ages are based on the distribution of the best estimate of intrauterine weights, whereas conventional birthweight charts are based on the weights of live births. The reported benefits of customised percentiles could therefore be attributed to either (1) the process of adjusting for maternal characteristics or (2) the incorporation of an intrauterine standard instead of a birthweight standard at younger gestational ages. Before concluding that the process of customising weight-for-gestational-age percentiles for maternal characteristics is beneficial, the separate contributions of each of these two methodological differences need to be understood.

Based on work previously performed by our group,[8] we had reason to believe that the overall improved prediction of mortality obtained with the use of customised birthweight charts may be due to methodological differences between intrauterine and conventional charts, rather than the process of adjusting for maternal characteristics. We therefore hypothesised that the improved prediction of perinatal morbidity and mortality previously demonstrated with the use of customised birthweight percentiles could also be obtained with a non-customised but intrauterine-based standard and that the regression-based adjustment for maternal characteristics may be an unnecessary step.

In this study, our first objective was to assess whether the improved prediction of perinatal mortality obtained through the use of a customised birthweight standard can also be obtained through the use of a non-customised but intrauterine-based standard. Our second objective was to quantify the extent to which the maternal characteristics in the customisation model are able to explain variability in birthweight to understand the amount of additional information that these variables provide to the prediction of optimal fetal weight.

Methods

Study population

The study population was drawn from singleton births ≥28 weeks of gestation in the Swedish Medical Birth Register between the years 1992–2001. The register contains information on 98–99% of births in Sweden, including stillbirths from 28 weeks of gestation.[9,10] The accuracy of the gestational ages, birthweights, and stillbirths in the register has previously been validated.[9,10] We excluded infants with congenital anomalies and infants with missing data on sex, birthweight, gestational age, or maternal covariates (height, pre-pregnancy weight, parity, age, or place of birth), leaving 81.5% of the original population. Further details on the Swedish Medical Birth Register and the final study sample are provided elsewhere.[3,8] The study was approved by the McGill Faculty of Medicine Institutional Review Board.

Calculation of customised and population percentiles

Customised birthweight percentiles were calculated according to previously published methods.[1,2] Briefly, an 'optimal' birthweight for 280 days of gestation was calculated based on covariates obtained from stepwise multiple regression (maternal height, pre-pregnancy BMI, ethnicity, parity, and fetal sex), then this weight was extrapolated to the optimal weight for the gestational age at birth using Hadlock's proportionality formula. This formula expresses 'optimal' predicted birthweight at earlier gestational ages as a proportion of the predicted weight at 280 days, using the fetal growth curve of Hadlock[7] to determine the trajectory through which the 280-day weight is reached. Customised percentiles were then calculated based on the discrepancy between optimal and actual birthweight.

Population-based percentiles were calculated in two alternative ways: (1) a sex- and gestational-week-specific birthweight reference based on the weight distribution of live births in this population[8] and (2) the intrauterine (ultrasound) estimated fetal-weight-for-gestational-age percentile chart published by Hadlock.[7] Although an intrauterine standard is available for Swedish populations,[11] Hadlock's standard was chosen for reasons of comparability since the proportionality formula in the commonly used and publicly available customised standard is derived from Hadlock's standard.[2] Small for gestational age (SGA) was defined as a birthweight below the 10th percentile, based on each of the birthweight, intrauterine weight, and customised standards (denoted as: SGA(birthweight), SGA(intrauterine) and SGA (customised), respectively).

Outcomes

The primary outcomes in this study were the occurrence of stillbirth and early neonatal death. Stillbirth was defined as a fetal death at 28 weeks of gestation or later and included both antepartum and intrapartum fetal deaths. Early neonatal mortality was defined as the death of a liveborn infant before 7 days of age.

Statistical analyses

The relative risks (RRs) of stillbirth among SGA infants as defined by each of the three standards were calculated with 95% confidence intervals (CI) using the infants classified as 'non-SGA' by the same standard as the referent group. Denominators in the calculations of risk were based on the number of continuing pregnancies at risk of stillbirth at each gestational age. The number of fetuses at risk (rather than live births) is the methodologically appropriate denominator in the study of stillbirths since all unborn fetuses are at risk for stillbirth, and

by definition, live births are no longer at risk of being stillborn.[12,13] Because Hadlock's intrauterine standard is not sex-specific,[7] a generalised linear model (binomial family, log link) was used to calculate sex-adjusted RRs for the intrauterine standard. RRs were calculated separately for term/post-term (≥37 weeks), mild preterm (34–36 weeks), and moderate-to-severe preterm (28–33 weeks) periods. Calculations were repeated using the outcome of early neonatal death. Fetuses at risk were also used as the denominator in the calculation of risk for early neonatal death[13] because all live fetuses are at risk of birth, and hence of neonatal death. Moreover, the risk factor of interest, intrauterine growth restriction, is known to lead to preterm birth (due to either obstetric intervention or, to a lesser degree, spontaneous birth).[14] Since preterm birth is a downstream effect of the risk factor, stratifying on birth status (by restricting to live births) would inappropriately adjust out the effects of poor fetal growth on mortality.

Multivariable linear regression was used to quantify the extent to which variables in the customisation model were able to explain variability in term birthweight (259–293 days, inclusive, similar to previously published models[1,2]), as established by the adjusted R^2. Since we were interested in understanding the contribution of maternal characteristics independent of the contributions of sex and gestational age (which are already 'customised' for in age- and sex-specific weight-for-gestational-age population standards), two separate models were built. The first contained only sex and gestational age at birth (linear and quadratic terms) as independent predictors, while the second additionally included the maternal characteristics (height, pre-pregnancy BMI, ethnicity, and parity). In both models, gestational age was centred at 280 days. The amount of variability explained by maternal characteristics alone was established as the incremental difference between the model with only sex and gestational age and the 'full' customisation model. Statistical analyses were conducted using Intercooled STATA 9.0 (Stata Corporation, College Station, TX, USA).

Results

Descriptive characteristics of the 782 303 infants and mothers in the final study population are presented in Table 1. The percentage of infants classified as SGA by each of the standards is shown in Table 2 according to gestational age at birth. While the percentage of infants classified as SGA by the population-based birthweight standard remained close to 10% across all gestational ages, the percentage of infants classified as SGA by the customised standard was much higher at early gestational ages, as high as 35% at 28–33 weeks. The percentages of SGA obtained with the intrauterine standard were very similar to those produced by the customised standard: 34% of infants born at 28–33 weeks were classified as SGA.

Table 1. Descriptive characteristics of the study population of 782 303 births in the Swedish Medical Birth Register, 1992–2001

Maternal characteristics	**Mean ± SD or *n* (%)**
Pre-pregnancy BMI (kg/m^2)	23.9 ± 4.0
Maternal height (cm)	166.3 ± 6.2
Maternal age (years)	28.9 ± 5.0
Parity (% nulliparous)	325 247 (41.6)
Country of birth (% Nordic*)	680 960 (87.1)
Fetus/infant characteristics	
Birthweight (g)	3566.1 ± 552.2
Gestational age at birth (weeks)	39.4 ± 1.7
Stillbirth	2354 (0.3)
Early neonatal death (<7 days)	815 (0.1)

*Nordic = Sweden, Norway, Denmark, Finland, and Iceland.

When the RRs of stillbirth were calculated for SGA infants classified by each of the three standards (Table 3), two trends were apparent. First, the RRs of stillbirth among SGA(customised) and SGA(intrauterine) were extremely similar overall and within each gestational age group, supporting our study hypothesis. Second, the difference between the RR(customised) or RR(intrauterine) and the RR(birthweight) varied by gestational age. At term ages, the relative risks produced by the different standards were fairly similar, but at preterm ages, the relative risks obtained from the intrauterine and customised standards were significantly higher than those obtained from the birthweight standard. The overall improved prediction of stillbirth obtained with the customised and intrauterine standards (RR = 6.1 for customised standard and RR = 6.2 for intrauterine standard) compared with the birthweight standard (RR = 3.8) was therefore driven primarily by improved classification of SGA at early gestational ages. The 10th percentile threshold at 30 weeks as established by each of the three standards is shown in Figure 1 in relation to the weights of stillbirths at this age. The 10th percentile of the

Table 2. Percentage of infants identified as SGA by the customised, birthweight, and intrauterine standards in 782 303 births in the Swedish Medical Birth Register, 1992–2001

	Birthweight SGA, *n* (%)	**Intrauterine SGA, *n* (%)**	**Customised SGA, *n* (%)**
Delivery 28–33 weeks (*n* = 8116)	801 (9.9)	2725 (33.6)	2840 (35.0)
Delivery 34–36 weeks (*n* = 28 472)	2821 (9.9)	4406 (15.5)	4595 (16.1)
Delivery ≥37 weeks (*n* = 745 715)	73 613 (9.9)	65 966 (8.9)	70 854 (9.5)
All ages (*n* = 782 303)	77 235 (9.9)	73 097 (9.3)	78 289 (10.0)

Table 3. RR of stillbirth among SGA infants, as established through customised, birthweight, and intrauterine standards in 782 303 births in the Swedish Medical Birth Register, 1992–2001

Gestational age	Standards	SGA (<10th percentile)		Non-SGA*		RR** (95% CI)
		At-risk population (*n*)	Stillbirths, *n* (risk per 1000)	At-risk population (*n*)	Stillbirths, *n* (risk per 1000)	
All ages	Birthweight	77 235	687 (8.9)	705 068	1667 (2.4)	3.8 (3.4–4.1)
	Intrauterine	73 097	906 (12.4)	709 206	1448 (2.0)	6.2 (5.7–6.7)
	Customised	78 289	952 (12.2)	704 014	1402 (2.0)	6.1 (5.6–6.7)
Delivery 28–33 weeks	Birthweight	77 235	144 (1.9)	705 068	469 (0.7)	2.8 (2.3–3.4)
	Intrauterine	73 097	337 (4.6)	709 206	276 (0.4)	12.2 (10.4–14.3)
	Customised	78 289	342 (4.4)	704 014	271 (0.4)	11.4 (9.7–13.3)
Delivery 34–36 weeks	Birthweight	76 434	152 (2.0)	697 753	310 (0.4)	4.5 (3.7–5.4)
	Intrauterine	70 372	200 (2.8)	703 815	262 (0.4)	7.9 (6.6–9.5)
	Customised	75 449	212 (2.8)	698 738	250 (0.4)	7.9 (6.5–9.4)
Delivery ≥37 weeks	Birthweight	73 613	391 (5.3)	672 102	888 (1.3)	4.0 (3.6–4.5)
	Intrauterine	65 966	369 (5.6)	679 749	910 (1.3)	4.2 (3.7–4.7)
	Customised	70 854	398 (5.6)	674 861	881 (1.3)	4.3 (3.8–4.8)

*Reference category.
**RR (intrauterine) adjusted for fetal sex.

birthweight standard was nearly 300 g lower than that of the intrauterine standard, while the 10th percentiles produced by the customised standard (shown as a distribution rug plot below the histogram since each infant's customised 10th percentile weight is different) were clustered around the value of the intrauterine 10th percentile.

Table 4 summarises the corresponding results for early neonatal mortality. As with stillbirth, RRs of early neonatal

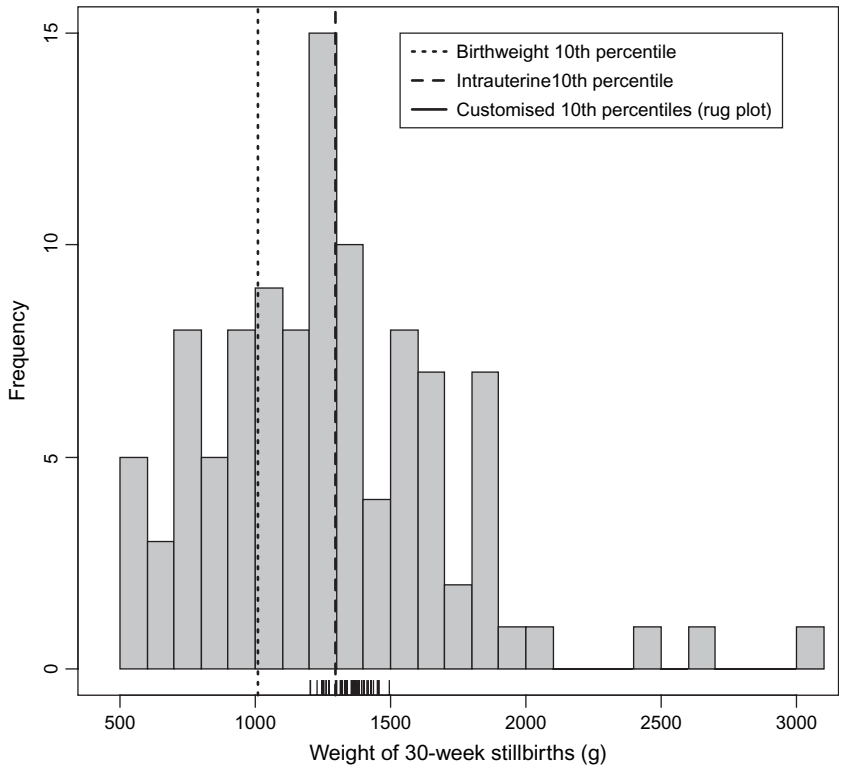


Figure 1. Tenth percentile weights of the birthweight (short dash), intrauterine (long dash), and customised standards (solid lines in rug plot below histogram) in relation to the weights of stillbirths at 30 weeks (histogram).

Table 4. RR of early neonatal death among SGA infants, as established through customised, birthweight, and intrauterine standards in 782 303 births in the Swedish Medical Birth Register, 1992–2001

Gestational age	Standards	SGA (<10th percentile)		Non-SGA*		RR** (95% CI)
		At-risk population (*n*)	Early neonatal deaths, *n* (risk per 1000)	At-risk population (*n*)	Early neonatal deaths, *n* (risk per 1000)	
All ages	Birthweight	77 235	229 (3.0)	705 068	586 (0.8)	3.6 (3.1–4.2)
	Intrauterine	73 097	297 (4.1)	709 206	518 (0.7)	5.9 (5.1–6.8)
	Customised	78 289	328 (4.2)	704 014	487 (0.7)	6.1 (5.3–7.0)
Delivery 28–33 weeks	Birthweight	77 235	43 (0.6)	705 068	182 (0.3)	2.2 (1.5–3.0)
	Intrauterine	73 097	100 (1.4)	709 206	125 (0.2)	8.3 (6.3–10.8)
	Customised	78 289	108 (1.4)	704 014	117 (0.2)	8.3 (6.4–10.8)
Delivery 34–36 weeks	Birthweight	76 434	59 (0.8)	697 753	127 (0.2)	4.2 (3.1–5.8)
	Intrauterine	70 372	71 (1.0)	703 815	115 (0.2)	6.5 (4.9–8.8)
	Customised	75 449	80 (1.1)	698 738	106 (0.2)	7.0 (5.2–9.3)
Delivery ≥37 weeks	Birthweight	73 613	127 (1.7)	672 102	277 (0.4)	4.2 (3.4–5.2)
	Intrauterine	65 966	126 (1.9)	679 749	278 (0.4)	4.9 (4.0–6.1)
	Customised	70 854	140 (2.0)	674 861	264 (0.4)	5.1 (4.1–6.2)

*Reference category.
**RR (intrauterine) adjusted for fetal sex.

mortality were similar for infants classified as SGA by the intrauterine standard and infants classified as SGA by the customised standard. Both were higher than the RR obtained with the conventional birthweight-for-gestational-age chart's classification of SGA. The overall improved prediction of early neonatal mortality obtained with the customised and intrauterine standards was again primarily derived from improved prediction at preterm gestational ages, with the three standards yielding comparable RRs at term.

Multivariable linear regression was used to compare explained variance in term birthweight (Table 5) between models with and without customisation for maternal characteristics. The regression model that included only gestational age and sex (Model 1) explained 17% of the variance in term birthweight. Once the sex and gestational age of the infant were known, information on maternal characteristics resulted in a modest improvement in the prediction of birthweight. Including maternal characteristics into the model explained an additional 7% of variance in birthweight (24% explained by the full model versus 17% explained by gestational age and sex alone).

Discussion

In this study, we have shown that a non-customised intrauterine weight standard has a similar ability to predict perinatal mortality as a customised birthweight standard. We conclude that the improved prediction of perinatal morbidity and mortality by customised birthweight percentiles compared with conventional birthweight percentiles is not derived from their adjustment for maternal characteristics (which is widely believed to be responsible for their apparent benefits[15]) but rather is derived from their use of Hadlock's intrauterine-based proportionality formula at preterm gestational ages. Maternal characteristics contributed little additional information to the customisation model compared with the information obtained from sex and gestational age at birth, explaining why customisation of population-based weight percentiles did little to further improve the prediction of mortality.

The validity of these conclusions is supported by the consistency of our results with previous reports on customised percentiles. The percentages of infants classified as SGA by the customised standard in this study were very similar to those reported by Groom *et al.*[16] (29.1% <34 weeks, 18.0% at 34–36[6] weeks, and 9.5% at ≥37 weeks) as well as to previous studies estimating the percentage of preterm live births classified as SGA by an intrauterine standard.[11,17–19] The increased RR of perinatal mortality among our study subjects classified as SGA by the customised standard was similar to results of previous studies of customised percentiles.[3,5] Unlike these earlier studies, however, we extended our evaluation of customised percentiles to additionally include a comparison with an intrauterine standard and found no statistically significant difference in the RR of stillbirth or early neonatal death between the customised and intrauterine standard. Conventional birthweight charts are biased at preterm gestational ages because of the association between fetal growth restriction and preterm birth,[11,20–24] and our results demonstrate that correction of this bias through the use of an intrauterine

Table 5. Explained variance in multiple linear regression models of birthweight among 688 529 term births (259–293 days of gestation, inclusive) in the Swedish Medical Birth Register 1992–2001

	Model 1 (gestational age and sex only)		Model 2 (full customisation model)	
	Coefficient	**95% CI**	**Coefficient**	**95% CI**
Sex (male)	121.1	119.0–123.2	120.6	118.6–122.6
Gestational age* (days)	23.1	22.9–23.2	22.2	22.1–22.4
Gestational age^2 (days)	−0.2	−0.2 to −0.2	−0.2	−0.2 to −0.2
Parity**				
Para 1			146.7	144.4–148.9
Para ≥2			177.5	174.9–180.2
Pre-pregnancy BMI (kg/m^2)***				
<18.5			−153.2	−159.1 to −147.3
25–29.9			107.1	104.7–109.6
≥30			175.9	172.1–179.6
Height (cm)******				
<160			−119.2	−122.3 to −116.0
>170			120.7	118.5–123.0
Ethnicity (non-Nordic)			−62.7	−65.8 to −59.7
Intercept	3557.2	3555.4–3558.9	3415.0	3412.7–3417.4
Adjusted R^2	0.17		0.24	

*Gestational age centred at 280 days.
**Reference category nullipara.
***Reference category 18.5–24.9 kg/m^2.
****Reference category 160–170 cm.

standard[19,25] improves identification of high-risk fetuses. Once this bias has been corrected, the process of customising for maternal characteristics does little to improve prediction of perinatal mortality. In the single customisation study that used a population standard based on intrauterine weights rather than birthweights, customisation was also found to have no benefits in the identification of growth-restricted stillbirths.[26] The large, population-based sample in our study provided sufficient statistical precision to detect differences and similarities between the customised, intrauterine, and birthweight standards in the prediction of perinatal mortality.

Although the proportion of variance in birthweight explained by our full customisation model was comparable to published values (e.g. $R^2 = 0.27$ in a customisation model for a French population[5]), the majority of the variance explained was due to the contributions of gestational age and sex. The absolute amount of variance explained by maternal characteristics was small (7%), which likely explains why the process of adjusting for maternal characteristics does not add meaningful information to distinguish truly growth-restricted infants from 'small-but-healthy' infants. The minor effect of adjusting for maternal characteristics is consistent with the results of our recent publication based on the same Swedish study sample.[8] We found that controlling for the increased percentage of preterm infants classified as SGA by the customised standard led to a large reduction in the relative odds of stillbirth (unadjusted OR = 7.8, OR adjusted for gestational age = 2.4). If adjustments for maternal characteristics were important in improving identification of stillbirth, such a large reduction would not have been seen. The small absolute amount of additional information provided by the maternal characteristics was a substantial proportion of the variance explained by the full customisation model, however. This result serves to highlight that a much better understanding of the physiological influences on fetal growth is likely needed before individualised risk prediction can be successful. Although we adjusted for similar maternal characteristics as those included in previous customisation models, it is possible that future customisation models with additional independent predictors may lead to improved prediction of birthweight.

The RRs presented in this paper should be interpreted with the limitations of the data in mind. As in previous studies, the gestational age of stillbirths was the age of birth, not the age of death, and it is unknown how much time elapsed between fetal demise and delivery. Likewise, the recorded weight of stillbirths was the weight at birth, not at the time of death. Any loss of weight between the time of death and time of delivery would be expected to overestimate the predictive value of weight-for-age charts.[27] Although these shortcomings may have introduced error to our estimates of risk, there is no reason to believe that such an error would have been

differential among the three standards (since all calculations used the same data on weights and gestational ages). It is therefore unlikely to affect the conclusions of our study, which focused on a comparison between intrauterine and customised standards rather than absolute effect.

Conclusions

Customised birthweight percentiles have already been recommended for clinical use by the UK's Royal College of Obstetricians and Gynaecologists,[6] and a recent editorial has called for American obstetricians to adopt their use as well.[15] The results of our study demonstrate that while the customisation for maternal factors does not impede the identification of high-risk infants, the process also provides little additional predictive benefit. The important contribution of past work by Gardosi *et al.*[1] appears to be not so much their regression model to 'customise' for maternal characteristics, but their recognition of the inappropriateness of using a birthweight-based standard at preterm ages.[25] Since data on maternal characteristics are often missing, even in high-quality databases such as the Swedish Medical Birth Register used in this study, a non-customised but intrauterine-based standard may be the most parsimonious and practical standard for the prediction of perinatal mortality in clinical practice.

Disclosure of interests

We have no conflicts of interest to declare.

Contribution to authorship

J.A.H. wrote the study protocol, performed the analyses, and wrote the first draft of the manuscript. X.Z., M.S.K., S.C., and R.W.P. all made substantial contributions to the study design, interpretation of results, and manuscript revisions. X.Z. established the study dataset (used in a previous manuscript *BJOG* 2007;114:474–77), obtained from S.C.

Details of ethics approval

This study received an expedited approval by the McGill University Faculty of Medicine Institutional Review Board on 15 November 2007 (A11-E38-07B).

Funding

J.A.H. is the recipient of a Doctoral Research Award from the Canadian Institutes of Health Research, R.W.P. is a Chercheur-Boursier of the Fonds de la recherche en santé du Québec. The Montreal Children's Hospital Research Institute receives operating funds from the Fonds de la recherche en santé du Québec. ■

References

1 Gardosi J, Chang A, Kalyan B, Sahota D, Symonds EM. Customised antenatal growth charts. *Lancet* 1992;339:283–7.
2 Gardosi J, Mongelli M, Wilcox M, Chang A. An adjustable fetal weight standard. *Ultrasound Obstet Gynecol* 1995;6:168–74.
3 Clausson B, Gardosi J, Francis A, Cnattingius S. Perinatal outcome in SGA births defined by customised versus population-based birthweight standards. *BJOG* 2001;108:830–4.
4 McCowan LME, Harding JA, Stewart AW. Customised birthweight centiles predict SGA pregnancies with perinatal morbidity. *BJOG* 2005; 112:1026–33.
5 Ego A, Subtil D, Grange G, Thiebaugeorges O, Senat MV, Vayssiere C, *et al.* Customized versus population-based birth weight standards for identifying growth restricted infants: a French multicenter study. *Am J Obstet Gynecol* 2006;194:1042–9.
6 Royal College of Obstetricians and Gynaecologists. *The Investigation and Management of the Small-for-Gestational-Age Fetus*. Guideline No. 31. London: RCOG, 2002.
7 Hadlock FP, Harrist RB, Martinez-Poyer J. In utero analysis of fetal growth: a sonographic weight standard. *Radiology* 1991;181: 129–33.
8 Zhang X, Platt RW, Cnattingius S, Joseph KS, Kramer MS. The use of customised versus population-based birthweight standards in predicting perinatal mortality. *BJOG* 2007;114:474–7.
9 Socialstyrelsens Epidemiologiska Centrum. *Utvärdering av det svenska Medicinska födelseregistret*. Stockholm, Sweden: Sweden, 2002; Article 2002-112-4 [www.socialstyrelsen.se/Publicerat/2002/2023/2002-112-4.htm]. Accessed 16 November 2007.
10 Cnattingius S, Ericson A, Gunnarskog J, Kallen B. A quality study of a medical birth registry. *Scand J Soc Med* 1990;118:143–8.
11 Marsal K, Persson PH, Larsen T, Lilja H, Selbing A, Sultan B. Intrauterine growth curves based on ultrasonically estimated foetal weights. *Acta Paediatr* 1996;85:843–8.
12 Yudkin PL, Wood L, Redman CWG. Risk of unexplained stillbirth at different gestational ages. *Lancet* 1987;1:1192–4.
13 Joseph KS. Incidence-based measures of birth, growth restriction, and death can free perinatal epidemiology from erroneous concepts of risk. *J Clin Epidemiol* 2004;57:889–97.
14 Morken N-H, Kallen K, Jacobsson B. Fetal growth and onset of delivery: a nationwide population-based study of preterm infants. *Am J Obstet Gynecol* 2006;195:154–61.
15 Resnik R. One size does not fit all. *Am J Obstet Gynecol* 2007;197: 221.
16 Groom KM, Poppe KK, North RA, McCowan LM. Small-for-gestational age infants classified by customized or population birthweight centiles: the impact of gestational age at delivery. *Am J Obstet Gynecol* 2007; 197:239. e1–5.
17 Zaw W, Gagnon R, da Silva O. The risks of adverse neonatal outcome among preterm small for gestational age infants according to neonatal versus fetal growth standards. *Pediatrics* 2003;111:1273–7.
18 Lackman F, Capewell V, Richardson B, daSilva O, Gagnon R. The risks of spontaneous preterm delivery and perinatal mortality in relation to size at birth according to fetal versus neonatal growth standards. *Am J Obstet Gynecol* 2001;184:946–53.
19 Bernstein IM, Mohs G, Rucquoi M, Badger GJ. Case for hybrid "fetal growth curves": a population-based estimation of normal fetal size across gestational age. *J Matern Fetal Med* 1996;5:124–7.
20 Weiner CP, Sabbagha RE, Vaisrub N, Depp R. A hypothetical model suggesting suboptimal intrauterine growth in infants delivered preterm. *Obstet Gynecol* 1985;65:323–5.
21 Ott WJ. Intrauterine growth retardation and preterm delivery. *Am J Obstet Gynecol* 1993;168:1710–5.

22 Secher NJ, Hansen K, Thomsen BL, Keiding N. Growth retardation in preterm infants. *Br J Obstet Gynaecol* 1987;66:463–71.
23 Fry AG, Bernstein IM, Badger GJ. Comparison of fetal growth estimates based on birth weight and ultrasound references. *J Matern Fetal Med* 2002;12:247–52.
24 Cooke RWI. Conventional birth weight standards obscure fetal growth restriction in preterm infants. *Arch Dis Child Fetal Neonatal Ed* 2007; 92:189–92.
25 Hutcheon JA, Platt RW. The missing data problem in birth weight percentiles and thresholds for "small-for-gestational-age". *Am J Epidemiol* 2008;167:786–92.
26 Lyon V, Howatson A, Khan K, Owen P. Unadjusted and customised weight centiles in the identification of growth restriction among stillborn infants. *BJOG* 2004;111:1460–3.
27 Chard T. Does the fetus lose weight in utero following infant death: a study in preterm infants. *BJOG* 2001;108:1113–15.

