Constructing Normalcy and Discrepancy Indexes for Birth Weight and Gestational Age Using a Threshold Regression Mixture Model

G. A. Whitmore[1], Guangyu Zhang[2], Mei-Ling Ting Lee[2,3]

1: McGill University, Montreal, and Ottawa Hospital Research Institute, Ottawa, Canada

2: Department of Epidemiology and Biostatistics, University of Maryland, College Park, MD, U.S.A.

3: National Cheng Kung University, Tainan, Taiwan

Abstract: Birth weight and gestational age are important measures of a newborn's intrinsic health, serving both as outcome measures and explanatory variables in health studies. The measures are highly correlated but occasionally inconsistent. We anticipate that health researchers and other scientists would be helped by summary indexes of birth weight and gestational age that give more precise indications of whether the birth outcome is healthy or not. We propose a pair of indexes that we refer to as the birth normalcy index or BNI and birth discrepancy index or BDI. Both indexes are simple functions of birth weight and gestational age and in logarithmic form are orthogonal by construction. The BNI gauges whether the birth weight and gestational age combination are in a normal range. The BDI gauges whether birth weight and gestational age are consistent. We present a three-component mixture model for BNI, with the components representing premature, at-risk and healthy births. The BNI distribution is derived from a stochastic model of fetal development proposed by Whitmore and Su (2007) and takes the form of a mixture of inverse Gaussian

distributions. We present a non-central t-distribution as a model for BDI. BNI and BDI are also well suited for making comparisons of birth outcomes in different reference populations. A simple z-score and t-score are proposed for such comparisons. The BNI and BDI distributions can be estimated for births in any reference population of interest using threshold regression.

Keywords: fetal development, healthy birth outcome, inverse Gaussian distribution, mixture model, stopping time, Wiener process

Author contact: Mei-Ling Ting Lee, Ph.D, Department of Epidemiology and Biostatistics, University of Maryland, College Park, MD, U.S.A. Email: mltlee@umd.edu

1. Introduction

Researchers in the field of perinatal health have long recognized the dual importance of birth weight and gestational age in judging the health of newborns. Many health and epidemiological studies make use of one or both measures, either as outcome measures or as key explanatory variables. National statistical systems use summaries of one or both measures in reporting on infant health and mortality. The World Health Organization (1995) recommends use of birth weight, adjusted by gestational age, to assess prenatal and postnatal health.

Medical concerns with unusually small (or large) birth weights and with preterm births have a long history and a huge literature. Conventionally, births with weights under 2,500 grams or larger than 4,500 grams, or terms under 37 weeks, are judged to be at higher risk of infant death, congenital malformations, mental retardation, and other physical and neurological impairments. Infants who are small or large for their gestational age (abbreviated SGA and LGA, respectively) are also at increased risk of neonatal morbidity and mortality, as well as health complications in adulthood. Definitions of these terms vary somewhat. A common definition of SGA is a birth below the 10th percentile weight for its gestational age. Symmetrically, LGA is a birth weight above the 90th percentile for its gestational age. Intrauterine growth restriction (IUGR) and macrosomia are also troublesome conditions. IUGR refers to a broad class of abnormal anatomical or functional disorders that manifest themselves in restricted fetal growth. It is important to recognize that only a subset of SGA infants are afflicted by IUGR. A sizable proportion of SGA infants are constitutionally small but are otherwise normal. Macrosomia refers to a neonate who is abnormally large and, often, unhealthily so. The implications of these birth conditions have an extensive literature. For SGA, for example, see Saenger, Czernichow, *et al* (2007) and Argente, Mehls and Barrios (2010). For LGA, see Kerényi, Tamás, Kivimäki *et al* (2009). For IUGR, see Resnik (2002), Garite, Clark and Thorp (2004), Haram, Svendsen and Myking (2007), and Chauban, Gupta, Hendix, and Berghella (2009). For macrosomia, see Stotland, Caughey, Breed, and Escobar (2004), Chauhan,

Grobman, Gherman, *et al* (2005), Zhang, Decker, Platt, and Kramer (2008), and Henriksen (2008). For an overview of these conditions for U.S. births, background studies and research, and access to available official data, refer to the website of the National Center for Health Statistics at www.cdc.gov/nchs/fastats/births.htm.. The inconsistency of birth weight and gestational age that is reflected in indicators such as SGA and LGA may arise because of abnormalities in fetal development or inaccurate measurements of gestational ages. Many researchers have attempted to address the problem of incorrect gestational ages. See, for example, Platt, Abrahamowicz, Kramer, *et al* (2001), Tentoni, Astolfi, Pasquale, and Zonta (2004), Parker and Schenker (2007), and Ananth (2007).

An early approach to modeling birth weights was taken by Wilcox and Russell (1983) who proposed a two-component mixture distribution for birth weight. One component consists of a 'predominant distribution' that has a normal density function and represents normal births. The second component is a 'residual distribution' with an unspecified form, representing low weight births. Taking a very different tack, Whitmore and Su (2007) consider a birth outcome to be determined by stopping a stochastic process representing the fetal development process. They discuss the joint modeling of birth weight and gestational age but set the full topic aside and focus on birth weight alone, making reference to common difficulties with the reliability of gestational age data and the challenges of joint modeling. One of the Whitmore and Su models for birth weight is a two-component mixture distribution in which the components correspond to 'normal' and 'abnormal' births. They assume a Wiener diffusion process as the underlying stochastic process for fetal development and, consequently, the two components of their mixture have inverse Gaussian distributions. The authors fit their models using a 4% systematic sample of U.S. live births in 2002; a total of 160,577 births. They provide interpretations of their findings and discuss practical implications of their work. These earlier models form the point of departure for the ideas and results we present here.

2. Constructing Orthogonal Birth Normalcy and Discrepancy Indexes

In this research, we convert the bivariate information in the birth weight and gestational age for a newborn into two indexes that are statistically orthogonal and therefore convey uncorrelated information about the birth outcome. We refer to these two indexes as the *birth normalcy index*, or BNI for short, and the *birth discrepancy index*, or BDI for short. The BNI roughly describes whether the infant's birth weight and gestational age, taken together, are abnormal. The BDI describes the extent to which birth weight and gestational age are inconsistent.

2.1 Definitions of the Indexes. Let W and A denote the birth weigh (in kilograms) and gestational age (in weeks) of a newborn. The formulas for our two indexes are as follows:

$$\text{BNI} = \sqrt{WW_A} \quad \text{and} \quad \text{BDI} = \sqrt{\frac{W}{W_A}}, \quad \text{where } W_A = \kappa A^{\alpha}. \tag{1}$$

Notation W_A represents an estimate of birth weight that is derived from the newborn's gestational age A, using the power function shown in (1). We will refer to W_A as the *imputed birth weight* because it can be used as a surrogate for the actual birth weight if we only know the infant's gestational age. As W_A is a re-expression of gestational age on a weight scale, BNI is essentially a geometric mean of birth weight and gestational age. Correspondingly, BDI is the square-root of the ratio of these two measures. The power function defining the imputed birth weight depends on two constants, κ and α. These are empirical constants derived from birth data for the reference population. The constants are chosen in such a manner that measure W_A mimics birth weight W. The rationale for choosing these particular mathematical forms for the indexes and imputed birth weight and the numerical determination of the two constants, κ and α, will emerge in the subsequent development. We note, however, that our choice of exponent α ensures that the logarithms of these indexes are uncorrelated and, hence, our claim that BNI and BDI are orthogonal.

In definition (1), we assume that the imputed birth weight W_A is a power function of gestational age A. Clinical evidence about the fetal development process supports this assumption of a power

function. The functional relationship between fetal weight and fetal age, as well as other fetal body measurements, has been the subject of much study. Tables of fetal weight percentiles at different fetal ages have been prepared to guide clinicians in this matter (see, for example, Doubilet, Benson, Nadel and Ringer, 1997). These tables show evidence of a power relationship between fetal weight and fetal age. If we use u and r to denote fetal weight and fetal age (to distinguish fetal measurements from those of the newborn), then the evidence suggests a relationship of the form $u = \kappa r^{\alpha}$, where κ is a proportionality constant and α is the age exponent. We have explored the best choice for exponent α by regressing $\ln(u)$ on $\ln(r)$ based on median fetal weights at different fetal ages for males and females combined, as reported in Doubilet, Benson, Nadel and Ringer (1997, p. 245). The fitted exponent varies somewhat with the range of fetal ages being considered. For fetal ages ranging from 25 to 43 weeks, the best-fitting exponent α is close to 3.5. For fetal ages over the narrower range of 35 to 39 weeks, α is close to 3. Thus, as the fetus approaches birth, the power relationship between its weight and age is roughly cubic. In the next section, we will establish an empirical estimate for α in power function (1) that is close to 3.

2.2 Setting the Empirical Constants in the Imputed Birth Weight. The empirical constants κ and α in the function $W_A = \kappa A^{\alpha}$ need to be specified for the reference population of births. We will choose these constants so the distribution of the imputed birth weight W_A imitates the distribution of birth weight W in the reference population. More specifically, we will choose $\ln(\kappa)$ and α as the solutions of the two moment equations that equate the sample means and standard deviations of $\ln(W)$ and $\ln(W_A)$, where $\ln(W_A) = \ln(\kappa) + \alpha \ln(A)$. To illustrate the procedure, we employ the same 4% systematic sample of US births in 2002 used by Whitmore and Su (2007). Birth weight in these data is measured in kilograms and gestational age in weeks. The estimates of $\ln(\kappa)$ and α that equate the means and standard deviations of W and W_A are -10.633 and 3.2314, respectively. The corresponding value of κ is 24.105×10^{-6}. With these estimates, we obtain the descriptive statistics for $\ln(W)$ and $\ln(W_A)$ presented in panel (a) of Figure 1. The means

and standard deviations are identical by construction. Comparing the first and third quartiles, as well as the minimum and maximum values, we see that the distribution of $\ln(W)$ is slightly stretched away from the center toward the tails, relative to $\ln(W_A)$. Kurtosis and skewness are also somewhat larger for $\ln(W)$ than its counterpart $\ln(W_A)$. Although not identical, the pairs of descriptive statistics for $\ln(W)$ and $\ln(W_A)$ match quite well. The fact that gestational age is only recorded to the nearest week has some influence on the descriptive statistics of W_A. To give a visual impression of the match of $\ln(W)$ and $\ln(W_A)$, panels (b) and (c) of Figure 1 show histograms of their respective distributions, plotted to the same scale. The histograms look well matched, except for the greater kurtosis of $\ln(W)$.

The BNI and BDI indexes may be expressed in the following log-linear form.

$$\ln \text{BNI} = \frac{1}{2}\left[\ln(W) + \ln(W_A)\right] \tag{2}$$

$$\ln \text{BDI} = \frac{1}{2}\left[\ln(W) - \ln(W_A)\right] \tag{3}$$

These expressions for $\ln(\text{BNI})$ and $\ln(\text{BDI})$ are an equally weighted sum and difference of $\ln(W)$ and $\ln(W_A)$, respectively. As we have chosen the empirical constant α so $\ln(W)$ and $\ln(W_A)$ have the same standard deviation, the construction ensures that $\ln(\text{BNI})$ and $\ln(\text{BDI})$ are uncorrelated and, hence, are orthogonal. The logarithms of the indexes are essentially first and second principal components of the logarithmic birth outcomes $\ln(W)$ and $\ln(W_A)$.

3. Statistical Modeling of BNI and BDI.

In the last section, we proposed a pair of orthogonal BNI and BDI indexes to capture the information in birth weight W and gestational age A. In this section, we build statistical models for BNI and BDI on the assumption that the joint birth outcome (W, A) occurs when a fetal development process is 'stopped' by reaching a birthing boundary. We show later that the models provide an excellent fit for US birth data and, presumably, for birth data from other reference populations.

3.1 Birth as a Stopping Event in a Fetal Development Process. We extend the Whitmore and Su birth-weight model to describe the joint outcome of birth weight W and gestational age A. We alter their notation somewhat to suit our purposes. In their formulation, fetal development is described by a two-dimensional process $\{U(r), D(r)\}$, where $U(r)$ denotes the fetal weight at fetal age r and $D(r)$ denotes an unspecified latent measure of fetal physiological development at age r. By definition, both measures are zero at conception. We use the term 'fetus' here for convenience, avoiding the more precise distinction between an embryo and fetus, with the latter being the designation after about 20 weeks of development. In their model, birth is a *stopping event* that is triggered when fetal physiological development first reaches a *birthing boundary*. They take the boundary to be a fixed development level b. In other words, birth occurs when the development process $D(r)$ reaches level b for the first time. This stopping event 'freezes' both the gestational age A and the birth weight W, where the latter is given by $W = U(A)$.

Whitmore and Su assumed that the latent fetal development process, expressed as a function of fetal weight, follows a Wiener diffusion process with mean parameter $\mu > 0$ and unit variance. The unit variance was justified by the fact that the fetal development process is latent and, hence, unobserved. By this construction, birth weight W is the first hitting time of a boundary b by a Wiener process and therefore has an inverse Gaussian distribution parameterized by μ and b. If fetal weight u and fetal age r were exactly related by a power function of the form $u = \kappa r^{\alpha}$, as we have just discussed, then the claim that the fetal development process follows a Wiener diffusion implies that both birth weight W and imputed birth weight $W_A = \kappa A^{\alpha}$ would have the same inverse Gaussian distribution. It seems reasonable, therefore, to take BNI for a given birth as a random draw from an inverse Gaussian distribution because BNI is a geometric mean of W and W_A, with variable W_A mimicking variable W. Our modeling of BNI starts with this assumption. We will show that the assumption is well supported by the data.

3.2 A Brief Overview of the Stopping Time Distribution. As the stopping time in our model has

an inverse Gaussian distribution, we give a brief overview of this distribution family for later use. An inverse Gaussian random variable S has the following probability density function (p.d.f.) and survival function (s.f.), denoted by f and $\overline{F}$, respectively.

$$f(s|b,\mu) = \frac{b}{\sqrt{2\pi s^3}} \exp\left[-\frac{(b-\mu s)^2}{2s}\right] \tag{4}$$

$$\overline{F}(s|b,\mu) = \Phi\left[\frac{(b-\mu s)}{\sqrt{s}}\right] - \exp(2b\mu)\Phi\left[-\frac{(b+\mu s)}{\sqrt{s}}\right] \tag{5}$$

Here $\Phi(\cdot)$ is the cumulative distribution function (c.d.f.) of the standard normal distribution. The mean (m), coefficient of variation (c), and mode of an inverse Gaussian distribution are related to the Wiener mean parameter $\mu > 0$ and boundary b as follows:

$$\mathrm{m} = \frac{b}{\mu}, \quad \mathrm{c} = \frac{1}{\sqrt{b\mu}}, \quad \text{mode} = \mathrm{m}\left[\left(1+\frac{9\mathrm{c}^4}{4}\right)^{1/2} - \frac{3\mathrm{c}^2}{2}\right] \tag{6}$$

An inverse Gaussian distribution is always right skewed so its mode is smaller than its mean. The skewness is less pronounced the smaller is the coefficient of variation. We will also use the following chi-square property of an inverse Gaussian random variable S (see Shuster 1968).

$$\frac{(b-\mu S)^2}{S} = \frac{1}{\mathrm{c}^2}\left(\sqrt{\frac{\mathrm{m}}{S}} - \sqrt{\frac{S}{\mathrm{m}}}\right)^2 \sim \chi_1^2 \tag{7}$$

Statistical theory shows that the square of a standard normal random variable has a chi-square distribution with one degree of freedom. In (7), we see that a similar standardization for an inverse Gaussian random variable gives the same result. In section 4, we will use this property to define a z-score for BNI.

3.3 BNI Model and Estimation. Births are not governed by a single stochastic process and birthing boundary. Abnormalities of various kinds and degrees can affect the fetal development

process, the birthing boundary, or both. This variety of possible development paths implies that a model for the distribution of BNI will be a statistical mixture with two or more components – one component for each type or category of birth experience. Each birth is unique but this view is not helpful from a statistical perspective. As a practical compromise, we seek that smallest number of categories that accurately captures the essential variety. The main component of the mixture will represent that majority of births in which the fetal development process and birth event follow a healthy natural course. This component is the 'predominant distribution' of Wilcox and Russell (1983) or the normal mixture component of Whitmore and Su (2007). Previous models have assumed that abnormal births constitute a residual component (Wilcox and Russell) or an abnormal second component of the mixture model (Whitmore and Su). We will depart from these previous models by assuming that the mixture distribution for BNI has three components that are associated with three categories or grades of births. Preliminary analysis shows that two components are not sufficient and, as we show later, three components provide a very good fit to U.S. birth data. In addition, the three components have very sensible interpretations in terms of the basic variety found in birth outcomes. We will use random variable J, with possible outcomes $j = 1, 2, 3$, to denote the category of a birth. The categories are defined solely by the fitted three-component mixture model. Because our classification of births is based on a fitted model, the categories are self-calibrating and do not adhere to any externally imposed definitions. We will give the categories a practical interpretation when we discuss the fitted model. For the moment, as simple labels, we refer to categories 1, 2, and 3 as *healthy*, *at-risk*, and *premature* births, respectively. These labels serve as convenient handles but oversimplify the underlying reality, as we will show.

The category of a birth has a probability distribution $p_j = Pr(J = j), j = 1, 2, 3$, where $p_1 + p_2 + p_3 = 1$. We assume that, whichever of these three categories applies to a fetus, the development path will follow a Wiener process and eventually hit a birthing barrier, producing the birth weight W at gestational age A. Each of these processes, however, will have its own mean

and boundary parameters, namely, parameters μ_j and b_j for categories $j = 1, 2, 3$. Letting $f_j(\text{BNI})$ denote the p.d.f. of BNI for birth category $j = 1, 2, 3$, it follows that the overall p.d.f. of BNI for our reference population is the following three-component mixture of inverse Gaussian density functions.

$$f(\text{BNI}) = p_1 f_1(\text{BNI}) + p_2 f_2(\text{BNI}) + p_3 f_3(\text{BNI}) \tag{8}$$

The preceding notation suppresses the dependence of the functions on their process and boundary parameters, as well as on covariates.

Panel (a) of Table 1 shows the fitted BNI mixture model for the 159,363 non-missing birth outcomes in the 4% sample of US births in 2002. The model parameters were estimated using the method of maximum likelihood, implemented with a numerical gradient routine in *Stata*. The likelihood function to be maximized is the mathematical product of the likelihood values calculated from (8). The table presents estimates of the mean m_j, coefficient of variation c_j, mode_j, and mixture probability p_j for each of the components $j = 1, 2, 3$ of the BNI mixture model. The table also shows the estimates of the boundary b_j and Wiener mean drift μ_j of the latent Wiener process for the three components. No covariates are taken into account in this implementation, but will be considered later. The fitted model can be interpreted as the estimated population distribution of BNI for US births in 2002.

The fitted model indicates that BNI has an average value of 3.433 for healthy births (category 1), a lower average at 2.933 for at-risk births (category 2), and the lowest mean at 1.641 for premature births (category 3). The coefficients of variation are 10%, 19% and 59%, respectively. The estimated modal values of BNI in each category are lower than their corresponding means, with a huge separation of the mode and mean for premature births. As one check on the fitted model, we note that the modal outcome for birth weight and gestational age in the data set are 3.4 kilograms and 39 weeks, respectively, with weight truncated to a tenth of a kilogram. This outcome is the

benchmark for healthy births. The BNI calculation for this modal outcome is

$$\text{BNI} = \sqrt{WW_A} = \sqrt{3.4(3.338)} = 3.37, \quad \text{where } W_A = \kappa A^{\alpha} = (24.105\times 10^{-6})(39^{3.2314}) = 3.338$$

This BNI number closely matches the estimated mode of 3.383 for healthy births (category 1) in panel (a) of Table 1. Finally, in the table, we see that the proportions of births falling in each category are estimated to be 0.775, 0.198 and 0.027, respectively.

A graphical assessment of the fit of the BNI mixture model (8) is shown in panel (a) of Figure 2. The figure plots the empirical cumulative distribution function (c.d.f.) against the fitted c.d.f. for the model. The fit is very good, with barely noticeable departures of the plot from the diagonal line. The maximum absolute difference in the functions is 0.0084. The plot has a slightly saw-toothed pattern because gestational age is only reported to the nearest week. The fitted mixture model takes no account of covariates. As covariates produce significant differences in mixture components of the BNI models as well as their mixture probabilities, the very small departures seen here are to be expected. We subsequently examine the effects of covariates on the BNI mixture model.

3.4 BDI Model and Estimation. The value of BDI for a birth captures inconsistencies between birth weight and gestational age. We have already noted for an idealized birthing process that fetal weight u and fetal age r would progress in an approximate power relationship, leading to the corresponding power relationship $W = \kappa A^{\alpha}$ for birth weight and gestational age. Some departure from this exact power relationship is expected. First, natural variability is expected in the relationship. Second, gestational age is prone to measurement error. Finally, medical anomalies in the fetal development process can cause fetal weight and fetal age to become misaligned in a variety of ways. Putting all of these sources of variability together, we posit a distribution for BDI that is, in essence, a noise distribution. We expect the distribution to be nearly symmetrical but

somewhat over-dispersed relative to a normal distribution. We have therefore chosen a non-central t-distribution as a model for BDI.

We have fitted a non-central t-distribution to BDI for the 4% sample of US births in 2002 using the method of maximum likelihood. The results appear in panel (b) of Table 1. The non-central t-distribution has three parameters, namely, a location or non-centrality parameter (ℓ), a scale parameter (σ), and its degrees of freedom (df). As the mean values of the logarithms of birth weight W and imputed birth weight W_A have been aligned purposely, we expect the non-centrality parameter ℓ of this t-distribution to be very close to 1. Indeed, its estimate is 1.0024 for this data set. The scale parameter σ describes the dispersion of BDI in the population of births. The estimate of σ for US births in 2002 is $.07561$, which indicates that most outcomes of BDI will lie within about 8% of its central location near 1. The estimated degrees of freedom of the fitted non-central t-distribution are 5 (precisely, 5.056). This number supports our presumption that BDI is over-dispersed.

A graphical assessment of the fit of a non-central t distribution to BDI for our data set is shown in panel (b) of Figure 2. As for the BNI model, the figure plots the empirical c.d.f. against the fitted c.d.f. for the model. Again, the fit is very good, with barely noticeable departures of the plot from the diagonal line. The maximum absolute difference in the functions is 0.0072.

4. Practical Applications of BNI and BDI

The practical value of the proposed indexes lies in their usefulness as both outcome measures and explanatory variables in perinatal research, for clinical assessments of neonatal health, and for statistical analysis, interpretation, and reporting on the condition of newborns. We begin in this section with the interpretation of the indexes and their relation to such conventional constructs as SGA, LGA and IUGR. We then discuss their use in comparing birth outcomes in different reference populations. Finally, we demonstrate how to use threshold regression to study the relation of the indexes to key covariates that define important subpopulations of interest.

4.1 Interpreting the Indexes. Our interpretation of the indexes starts with BNI. The partial probability density functions $p_j f_j(\text{BNI})$ of the BNI mixture model, corresponding to birth categories $j = 1, 2, 3$, are plotted in panel (a) of Figure 3. The parameter estimates for these densities are presented in panel (a) of Table 1, including estimates of the boundary b_j and mean drift μ_j of the fetal development process for $j = 1, 2, 3$. Recall that the fetal development process for each category of birth has a unit variance by definition. We note that the boundary value and mean drift both decline as we move from the healthy to premature birth categories. These patterns produce the falling mean values and rising coefficients of variation seen in panel (a) of Table 1.

The dominant density function, representing 77% of births, corresponds to the healthy category ($j = 1$) and has the appearance of a normal density. The secondary density function corresponds to the at-risk category ($j = 2$) and represents 20% of births. This secondary density overlaps substantially with that of healthy births. With this observation, it is the moment to explain what the label 'at risk' encompasses. Recall that each category of birth arises from a separate fetal development process and birthing boundary. The at-risk category is one that has a shallower drift and lower birthing boundary than healthy births. This category covers a range of risky scenarios, from newborns at risk of being moderately premature to newborns of normal weight and gestational age who have other physiological problems that have manifested themselves in a more erratic fetal development trajectory. The latter may include significant congenital malformations, chromosomal damage, and other prenatal anomalies that place them at risk for later health problems. The minor density function in panel (a) of Figure 3 corresponds to the premature birth category ($j = 3$) and accounts for about 3% of births. It clearly represents moderately to severely premature newborns. The underlying development process for these infants has a boundary closer to zero than for at-risk births and a very shallow mean drift. These birth outcomes are therefore relatively volatile. This last fact is confirmed by the huge coefficient of variation for the premature category.

The mixture probabilities $p_j, j = 1, 2, 3$, may be considered as prior probabilities in the Bayesian

sense. They represent the chances that a birth will be in each of these categories when we have no knowledge of the infant's weight or gestational age. Bayes theorem, in conjunction with mixture model (8), can then be used to calculate the posterior probability that a birth lies in a given category, conditional on its BNI value, as follows.

$$p_j(\text{BNI}) = \frac{p_j f_j(\text{BNI})}{f(\text{BNI})}, \quad j = 1, 2, 3. \tag{9}$$

We have plotted these posterior probabilities against the birth normalcy index for the 4% sample of US births for 2002. The plot appears in panel (b) of Figure 3. The plot is revealing. At any value of BNI, the three probabilities add to 1, i.e., $p_1(\text{BNI}) + p_2(\text{BNI}) + p_3(\text{BNI}) = 1$. The major crossing points of the probability curves are at BNI values of 1.77 for the premature and at-risk curves, 2.80 for the healthy and at-risk curves, and a further crossing of the healthy and at-risk curves at 4.70. For BNI values below 1.77, a premature birth is more probable than an at-risk birth. At BNI values from 1.77 to 2.80, an at-risk birth is more probable than a healthy birth. These two breakeven points bracket births that range from being severely premature to mildly premature. In the BNI range from 2.80 to 4.70, a birth is most likely to be from the healthy category. Finally, the probability of an at-risk birth overtakes that of a healthy birth above a BNI of 4.70. Thus, above this highest breakeven point, there is growing risk that a birth is macrosomic. Another insight offered by these probability plots is that the category of a birth cannot be known with certainty from the BNI value alone, except possibly when the BNI is extremely low. A birth has a moderate chance (.07) of being at-risk, even when the probability of a healthy birth is at its peak (.93), which occurs at a BNI of 3.64.

Turning to the interpretation of BDI, we have already noted that it is measuring the inconsistency of birth weight and gestational age. Low and high percentiles (say, the 10th and 90th percentiles) of BDI may therefore be taken as refined indicators of SGA and LGA infants. The measures are refined in the sense that BDI is orthogonal to BNI. The BDI percentiles do not have

to *fix* the value of gestational age A in order to declare a misalignment of birth weight for the gestational age. The two percentiles of BDI used for declaring SGA and LGA births are a pair of numbers and not bands of percentiles that vary over gestational age. For US births in 2002, for example, the fitted non-central t-distribution in panel (b) of Table 1 has 10th and 90th percentiles of BDI equal to 0.891 and 1.114. The percentiles are centered slightly away from 1 because the estimated location parameter (ℓ) is 1.0024. For example, consider two births with outcomes (W, A) of $(2.309, 40)$ and $(1.497, 35)$. Both have the same BDI (0.798) and therefore exhibit equal degrees of discrepancy between birth weight and gestational age. As both births lie below the 10th percentile for BDI, they can be declared as SGA.

As a final note on interpreting the indexes, we note that if birth weight and gestational age are perfectly aligned, i.e., $W = W_A$, then BDI is exactly 1 and BNI can be interpreted directly as a birth weight in kilograms. The preceding breakeven points in panel (b) of Figure 3 can then be read off as birth weights.

4.2 Comparing Birth Outcomes in Different Reference Populations. We sometimes wish to judge the relative healthiness of births in different reference populations or subpopulations, such as in different ethnic subpopulations within a country (e.g., Hispanic and black populations in the U.S.A.) or in international comparisons across countries (e.g., Bangladesh, India, Pakistan and Sri Lanka). The BNI mean value for the healthy birth category (m_1) serves as a *norm* or benchmark for births in the reference population. Whitmore and Su (2007) propose a z-score for birth weights that serves this purpose. We extend their z-score to BNI as follows:

$$\text{z-score} = \frac{1}{\mathrm{c}_1}\left(\sqrt{\frac{\mathrm{BNI}}{\mathrm{m}_1}} - \sqrt{\frac{\mathrm{m}_1}{\mathrm{BNI}}}\right) \tag{10}$$

The healthy birth p.d.f. $f_1(\mathrm{BNI})$ in the BNI mixture model (8) is assumed to follow an inverse Gaussian distribution and the assumption is supported by the fact that the mixture model fits actual birth data very well. With plug-in estimates for parameters, the formula for the inverse Gaussian

s.f. in (5) can be used to make probability statements about the z-score. It also follows from (7) that the square of the z-score is an exact chi-square random variable with one degree of freedom. As the next illustration shows, the inverse Gaussian distribution can be used to calculate a one-sided P-value for a BNI value, whereas the chi-square distribution can be used to compute a two-sided P-value.

To illustrate the z-score for an international comparison, consider two countries A and B. The mean m_1 and coefficient of variation c_1 for healthy births in the two countries (derived from fitted mixture models for their national birth data) are, say, 3.30 and 0.10 for country A and 3.50 and 0.10 for country B, respectively. Consider a birth in each country that happens to have the same BNI value of 2.892. In the U.S., a birth weight of 2.309 kilograms and a gestational age of 40 weeks, for instance, would produce this BNI value. Computing the z-score for a BNI of 2.892, using the formula in (10), gives scores of -1.321 and -1.911 in countries A and B, respectively. One-sided P-values for these z-scores, calculated from (5), are .102 and .031, respectively. Two-sided P-values for these z-scores can be calculated from the χ^2_1 distribution. Squaring each z-score gives $(-1.321)^2 = 1.744$ and $(-1.911)^2 = 3.652$ for countries A and B. Reference to a computer routine for the χ^2_1 distribution gives two-sided P-values of .187 and .056, respectively. Note that the two-sided P-values are not exactly double their one-sided counterparts because of the slight skewness of an inverse Gaussian distribution. The implication of the one-sided P-values is that a healthy birth with a BNI of 2.892 or lower will arise about once in 10 healthy births in country A but only about once in 32 healthy births in country B. The birth outcome is therefore more outlying in country B than country A. The fact that the z-scores are negative indicate that both births are below the norms (3.30 and 3.50) for healthy births in their respective national populations.

Based on our finding that the BDI follows a non-central t-distribution, we can give the BDI a corresponding t-score, as follows:

$$\text{t-score} = \frac{\text{BDI} - \ell}{\sigma} \tag{11}$$

For any birth, the t-score follows a t-distribution with df degrees of freedom. Computer routines and tables for t distributions are widely available so again it is a simple matter to make probability statements about a t-score for any reference population or subpopulation. For example, consider a birth in the US in 2002 that has a BDI of 0.798. Panel (b) of Table 1 gives respective estimates for ℓ, σ and df of 1.0024, .07561 and 5.056. It follows therefore that the t-score $= (.798 - 1.0024)/.07561 = -2.703$. A computer routine for the t-distribution gives a probability of .021 of obtaining this or a smaller BDI value in a randomly chosen American birth. The birth is an SGA candidate.

4.3 Accounting for Covariate Effects on BNI and BDI. The parameters of the BNI mixture model (8) or the non-central t-distribution for BDI can be made to depend on selected covariates and then estimated using regression methods. We will demonstrate the regression analysis for BNI only. For this case demonstration, we again employ the 4% systematic sample of US births in 2002. We have selected the following three binary covariates: (1) *Under 20*: Age of mother, coded 1 if the mother is under 20 years and 0 otherwise; (2) *White Race*: Race of mother, coded 1 if white and 0 otherwise; and (3) *Vaginal Delivery*: Delivery method, coded 1 if vaginal without previous C-section and 0 otherwise. As the BNI model involves the inverse Gaussian distribution, we use threshold regression as described in Lee and Whitmore (2006, 2010). Threshold regression output for the BNI model with these covariates appears in Table 2. The fitted model uses the preceding covariates for each parameter, with identity link functions for the mean parameters, logarithmic link functions for the coefficients of variation, and logit link functions for the mixture probabilities. Each regression function is a simple linear combination of the covariates. No account has been taken of possible covariate interactions. The selection of covariates and the forms of the regression functions have been kept simple so the demonstration is not too elaborate. We are aware

that the regression model can be substantially refined.

We now comment briefly on the results in Table 2. The regression analysis involves 24 regression coefficients for the three covariates. Because of the large number of parameters being estimated, we take a regression effect to be significant only if its P-value is under .002. The parameter p_2 does not appear in the regression table because it depends on p_1 and p_3 through the identity $p_2 = 1 - p_1 - p_3$.

1. For premature births, the only significant regression coefficient ($P < .002$) for the mean m_3 and coefficient of variation c_3 is the positive effect of *Delivery* on the coefficient of variation. All of the covariates have significant influences on the probability of a premature birth p_3. Being a teen-aged mother raises the probability of a premature birth while being a white mother and having a vaginal delivery reduces the probability.

2. For births in the at-risk and healthy categories ($j = 2$ and $j = 1$, respectively), we see that almost all covariates affect the means m_2 and m_1 and coefficients of variation c_2 and c_1 in a significant way. The exceptions are *Under 20* and *White Race* for c_2 and *Vaginal Delivery* for m_1. The general tendency for at-risk and healthy births is for their BNI means to fall and their coefficients of variation to increase when the mother is a teenager. The effects are in the opposite direction for a mother being white or having a vaginal delivery, with the means and coefficients of variation rising and falling respectively for at-risk and healthy births. None of the covariates affects the probability of a healthy birth p_1. This finding is interesting because it says that the chance of a healthy birth is invariant but the location and relative variability of the BNI distribution for healthy births are not.

3. The mean value of BNI for healthy births (m_1) ranges from 3.29 to 3.46 over the eight possible configurations of *Under 20*, *White Race*, and *Vaginal Delivery* in the sample. Correspondingly, p_1 ranges from 0.73 to 0.83.

5. Concluding Discussion

An analogy can be drawn between our proposed indexes BNI and BDI and the body mass index BMI. The BMI measures the consistency of body weight and body height. Unusually small or large values of BMI indicate misalignments of weight and height, as occurs in obesity for example. It follows, therefore, that our BDI is the counterpart of BMI in that it measures the extent to which birth weight and gestational age are inconsistent. The exponent of 2 that is used for height in the BMI is the analog of our near-cubic α exponent for gestational age A in (1). Interestingly, to our knowledge, no analog of the BNI has been proposed for body weight and height, although our line of development presumably could be used to derive a body mass normalcy index.

The BNI and BDI are very suitable measures for research studies that wish to capture the joint predictive effects of birth weight and gestational age, especially where the research involves, say, regression models with birth weight and gestational age as explanatory variables. The fact that the logarithms of the indexes are uncorrelated implies that their regression effects will not be diluted or distorted by multicollinearity and would be measuring independent explanatory contributions.

Our case demonstration has focused on a 4% systematic sample of US births in 2002. In principle, our results are easily extended to all births for that year or even for multiple years. We have reserved the application of our methods to larger data sets for a subsequent research study of time trends in US birth patterns.

Acknowledgments The authors thank the National Center for Health Statistics for making their birth data available for researchers in a convenient format, together with ample supporting documentation. This research is supported in part by NIH Grant OH008649 (Lee).

Cited References

1. Argente J, Mehls O, Barrios V (2010). Growth and body composition in very young SGA children, *Pediatr Nephrol*, **25**, 679-685.

2. Ananth CV (2007). Menstrual versus clinical estimate of gestational age dating in the United States: temporal trends and variability in indices of perinatal outcomes, *Paediatric and Perinatal Epidemiology*, **21** (Suppl. 2), 22-30.

3. Chauhan SP, Grobman WA, Gherman RA *et al* (2005). Suspicion and treatment of the macrosomic fetus: A review, *Am J Obstet Gynecol*, **193**, 332-346.

4. Chauhan SP, Gupta LM, Hendix NW, Berghella V (2009). Intrauterine growth restriction: comparison of American College of Obstetricians and Gynecologists practice bulletin with other national guidelines, *Am J Obstet Gynecol*, **200**, 409.e1-409.e6.

5. Doubilet PM, Benson CB, Nadel AS, Ringer SA (1997). Improved birth weight table for neonates developed from gestations dated by early ultrasonography, *Journal of Ultrasound Medicine*, **16**, 241-249.

6. Garite TJ, Clark R, Thorp JA (2004). Intrauterine growth restriction increases morbidity and mortality among premature neonates, *Am J Obstet Gynecol*, **191(2)**:481-487.

7. Haram K, Svendsen E, Myking O (2007). Growth restriction: Etiology, maternal and neonatal outcome: A review, *Current Women's Health Reviews*, **3**, 145-160.

8. Henriksen T (2008). The macrosomic fetus: a challenge in current obstetrics, *Acta Obstetrica et Gynecologica*, **87**, 134-145.

9. Keréyi Z, Tamás G, Kivimäki M *et al* (2009). Maternal glycemia and risk of large gestational age babies in a population-based screening, *Diabetes Care*, To appear.

10. Lee M-LT, Whitmore GA (2006). Threshold regression for survival analysis: Modeling event times by a stochastic process reaching a boundary, *Statistical Science*, **21**, 501-513.

11. Lee M-LT, Whitmore GA (2010). Proportional hazards and threshold regression: their theoretical and practical connections, *Lifetime Data Analysis*, **16**, 196214.

12. Parker JD and Schenker N (2007). Multiple imputation for national public-use datasets and its possible application for gestational age in United States Natality files, *Paediatric and Perinatal Epidemiology*, **21**(Suppl. 2), 97-105

13. Platt RW, Abrahamowicz M, Kramer MS, Joseph KS, Mery L, Blondel B, Breart G, Wen SW (2001). Detecting and eliminating erroneous gestational ages: a normal mixture model, *Statistics in Medicine*, **20**: 3491-3503

14. Resnik R (2002). Intrauterine growth restriction, *Obstetrics and Gynecology*, **99**, 490-496.

15. Saenger P, Czernichow P, Hughes I, Reiter EO (2007). Small for gestational age: Short stature and beyond, *Endocr. Rev.*, **28**, 219-251.

16. Shuster JJ (1968). On the inverse Gaussian distribution function. *J. Amer. Statist. Ass.*, **63**, 1514-1516.

17. Stotland NE, Caughey AB, Breed EM, and Escobar GJ (2004). Risk factors and obstetric complications associated with macrosomia, *International Journal of Gynecology & Obstetrics*. **87(3)**:220-226.

18. Tentoni S, Astolfi P, Pasquale AD, Zonta LA (2004). Birthweight by gestational age in preterm babies according to a Gaussian mixture model, *International Journal of Obstetrics and Gynaecology*, **111**, pp. 31-37

19. World Health Organization Expert Committee on the Use and Interpretation of Anthropometry (1995). Physical status: the use and interpretation of anthropometry, Geneva: World Health Organization.

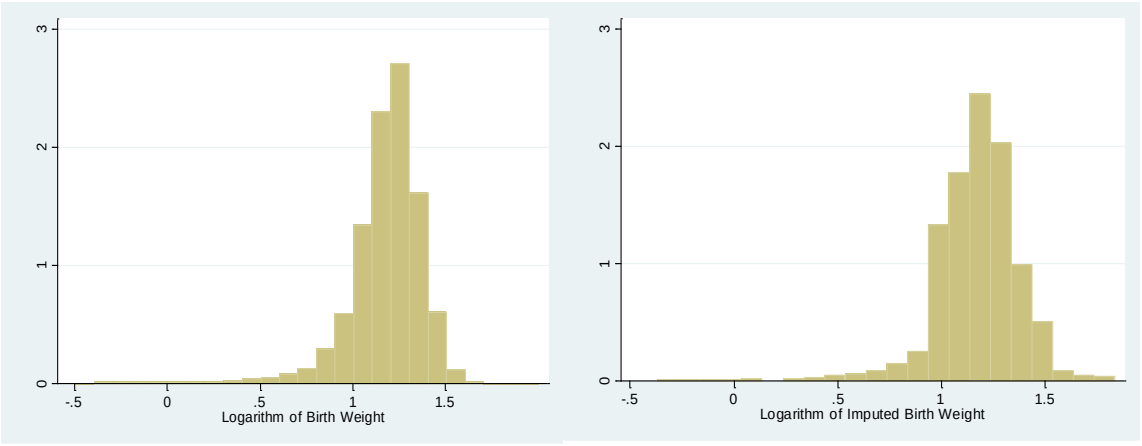
20. Whitmore GA, Su Y (2007). Modeling low birth weights using threshold regression: Results for U.S. birth data, *Lifetime Data Analysis*, **13**, 161-190.

21. Wilcox AJ, Russell IT. (1983). Birthweight and perinatal mortality. I. On the frequency distribution of birthweight, *International Journal of Epidemiology*, **12**, 314-318.

22. Zhang X, Decker A, Platt RW, Kramer MS (2008). How big is too big? The perinatal consequences of fetal macrosomia, *Am J Obstet Gynecol*, **198(5)**: 603-4; discussion e1-5.

(a)

Variable	Births	Mean	S. D.	1st Q.	Md.	3rd Q.	Skew.	Kurt.	Min.	Max.
$\ln(W)$	159,363	1.171	.2355	1.099	1.207	1.303	-3.344	23.74	-1.483	1.946
$\ln(W_A)$	159,363	1.171	.2355	1.121	1.205	1.287	-2.765	18.95	-1.478	1.808

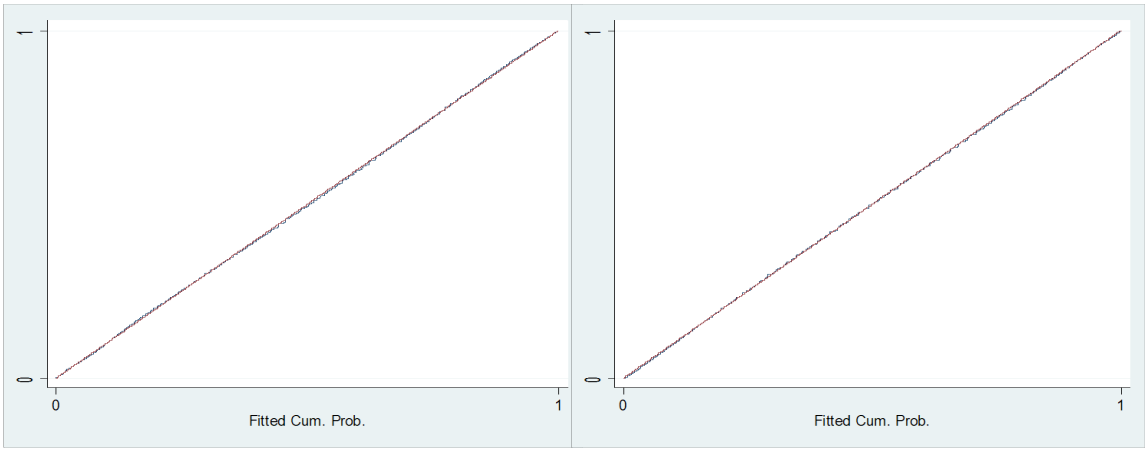


(b) **(c)**

Figure 1: Panel (a): Descriptive statistics for the logarithms of birth weight $\ln(W)$ and imputed birth weight $\ln(W_A) = \ln(\kappa) + \alpha \ln(A)$, where the constants κ and α are chosen so the two variables have the same mean and standard deviation. Panels (b) and (c): Histograms for $\ln(W)$ and $\ln(W_a)$. The histograms and statistics are based on a 4% sample of US births in 2002, excluding missing values. Birth weight is in kilograms and gestational age in weeks.

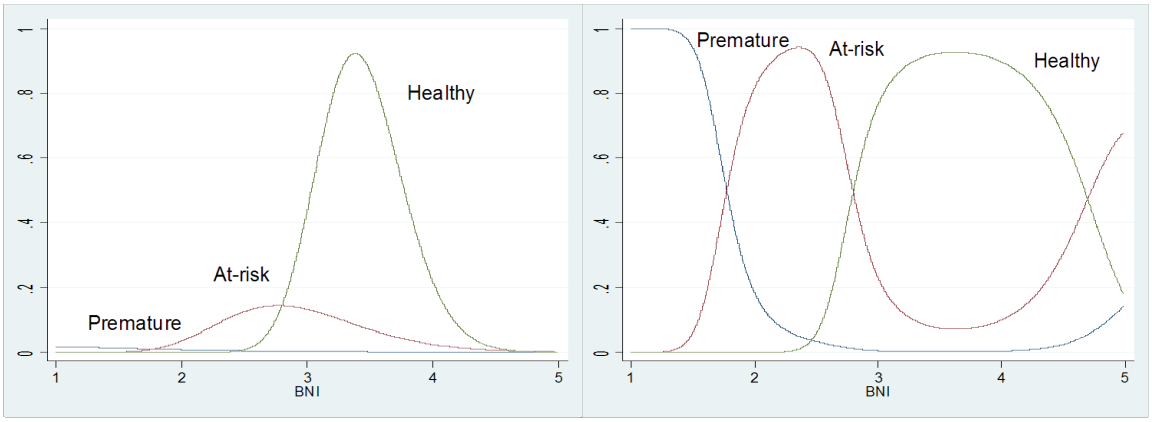
	Birth Category j		
Parameter	$j = 1$: Healthy	$j = 2$: At-risk	$j = 3$: Premature
(a) Estimated Inverse Gaussian Mixture Model for BNI for the US Population			
Mean (m_j)	3.433	2.933	1.641
Coef. of Var. (c_j)	.0987	.1949	.5886
Mode ($mode_j$)	3.383	2.767	0.843
Probability (p_j)	.7747	.1982	.0271
Boundary (b_j)	18.77	8.78	2.18
Mean drift (μ_j)	5.47	3.00	1.33
(b) Estimated Non-central t-distribution Model for BDI for the US Population			
Parameter			
Location (ℓ)	1.0024		
Scale (σ)	.07561		
Degrees of Freedom (df)	5.056		

Table 1: (a) Fitted three-component mixture model for BNI for the US population in 2002, based on a sample of 159,353 births. (b) Fitted non-central t-distribution for BDI for the US population in 2002.



(a) **(b)**

Figure 2: Plots of the empirical c.d.f. against the fitted c.d.f. as an assessment of goodness of fit. Panel (a): The three-component inverse Gaussian mixture model (8) for BNI. Panel (b): The non-central t-distribution model for BDI. Differences between the empirical and fitted c.d.f.s are almost imperceptible in both plots.



(a) **(b)**

Figure 3: Panel (a): Partial probability density functions $p_j f_j(\text{BNI})$ of the BNI mixture model, corresponding to birth categories $j = 1$ (healthy), $j = 2$ (at-risk) and $j = 3$ (premature). Panel (b): Probability of a given category of birth $p_j(\text{BNI})$ at each value of BNI, for $j = 1, 2, 3$.

Birth Category	$j = 1$: Healthy		$j = 2$: At-risk		$j = 3$: Premature	
	Estimate	P-value	Estimate	P-value	Estimate	P-value
m_j						
Under 20	-.0514	.000	-.1833	.000	-.2102	.028
White Race	.1119	.000	.2122	.000	.1986	.005
Vaginal Delivery	.0065	.080	.1597	.000	-.0448	.542
Constant	3.3417	.000	2.6722	.000	1.4706	.000
$\ln(\text{c}_j)$						
Under 20	.1021	.000	.0461	.219	-.1026	.155
White Race	-.0850	.000	-.0686	.003	-.0373	.396
Vaginal Delivery	-.0810	.000	-.1076	.000	.3580	.000
Constant	-2.2038	.000	-1.5116	.000	-.7134	.000
$\text{logit}(p_j)$						
Under 20	.2783	.018			.4590	.001
White Race	-.1148	.112			-.5474	.000
Vaginal Delivery	.1675	.011			-1.1871	.000
Constant	1.3087	.000			-1.0255	.000

Table 2: Fitted threshold regression model for BNI for the US population in 2002