Lifetime Data Anal (2007) 13:161–190
DOI 10.1007/s10985-006-9032-y

Modeling low birth weights using threshold regression: results for U. S. birth data

G. A. Whitmore · Yi Su

Received: 4 May 2006 / Accpeted: 28 December 2006 / Published online: 8 February 2007

Abstract Babies born live under 2,500 g or with a gestational age under 37 weeks are often inadequately developed and have elevated risks of infant mortality, congenital malformations, mental retardation, and other physical and neurological impairments. In this paper, we model birth weight as a first hitting time (FHT) of a birthing boundary in a Wiener process representing fetal development. We associate the parameters of the process and boundary with covariates describing maternal characteristics and the birthing environment using a relatively new regression methodology called *threshold regression*. Two FHT models for birth weight are developed. One is a mixture model and the other a competing risks model. These models are tested in a case demonstration using a 4%-systematic sample of the more than four million live births in the United States in 2002. An extensive data set for these births was provided by the National Center for Health Statistics. The focus of this paper is on the conceptual framework, models and methodology. A full empirical study is deferred to a later occasion.

Keywords Bayes analysis · Birth data · Birth weight · Competing risks · Covariates · Fetal development · First hitting time · Gestational age · Health statistics · Inverse Gaussian distribution · Low birth weight · Mixture model · Model checking · Premature birth · Preterm birth · Statistical inference · Statistical model · Subsampling · Threshold regression · Wiener stochastic process · z-score

G. A. Whitmore (✉) · Yi Su
Desautels Faculty of Management, McGill University, 1001 Sherbrooke Street West, Montreal, QC Canada H3A 1G5
e-mail: george.whitmore@mcgill.ca

1 Introduction

The maturity of newborns is often judged by their birth weight and gestational age (age at birth measured from conception). Low weight and preterm babies are found to have elevated risks of infant mortality, congenital malformations, mental retardation, and other physical and neurological impairments. Conventional cutoffs for birth weight and gestational age are babies born live under 2,500 g or under 37 weeks, respectively. Medical concerns with low weight and preterm births have a long history and a huge literature. A study by the National Center for Health Statistics (NCHS 1980), for example, provided a statistical overview of the problem three decades ago. Yet, that study report seems to have been written yesterday when judged by the slow progress that has been made in dealing with the medical challenges associated with these kinds of births. The effects of these births are better understood today but progress has been slow in discovering and tackling their root causes. For an overview of the current situation with respect to U.S. births, available official data, and background studies and research on this topic, the reader is referred to NCHS (2003) and the 125 references cited there. For a perceptive review of the measurement issues related to the course and outcome of pregnancy, the reader is referred to Savitz et al. (2002).

In this paper, we study a 4%-systematic sample of the more than four million live births in the United States in 2002. We model birth weight as a first hitting time of a birthing boundary in a Wiener process representing fetal development. We associate the parameters of the boundary and process with covariates representing characteristics of the mother and the birthing environment. We employ a regression methodology called *threshold regression* to link the parameters to these covariates. The word 'threshold' here refers to the birthing boundary that defines the first hitting time. This usage is not to be confused with applications in fields where phenomena have a threshold level of response, as in toxicology for example.

The study of regression structures for first hitting times in Wiener processes is not new (see, for example, Whitmore 1983). It takes on its more modern form in Lee et al. (2000) and Lee et al. (2004). Lee and Whitmore (2004, forthcoming) contain reviews of threshold regression for first hitting times.

We develop two first hitting time models for birth weights here. These are tested in a regression case demonstration using the 4%-systematic sample of U.S. live births in 2002. The focus of this paper is on the conceptual framework, the models and their associated methods. A full empirical study of the data set is deferred to later research. The data are so extensive and involve so many health and medical technicalities that the requisite research and reporting will require several separate publications to other audiences.

2 Demographic and clinical variables associated with low birth weight

Final microdata for U.S. live births in 2002 are available from the NCHS in a CD format (NCHS 2003). The variables covered by this NCHS data set are

traditional demographic and clinical variables that have been used for decades in statistical reporting by the NCHS. The following partial listing illustrates the range and nature of variables covered by the 2002 data set.

Maternal demographic characteristics State and county of residence, place of birth; numbers of live births now living and dead, interval since last live birth, age, live-birth order, race, Hispanic origin, marital status, and educational attainment.
Paternal demographic characteristics Age, race, Hispanic origin.
Maternal medical characteristics Medical risk factors, weight gain during pregnancy, tobacco, and alcohol use.
Medical care utilization by pregnant women Prenatal care, obstetric procedures, complications of labor and/or delivery, attendant at birth, and method of delivery.
Infant characteristics State and county of birth, sex, date of birth, gestational duration, birth weight, Apgar score, abnormal conditions, congenital anomalies and multiple births.

The listing reminds us of the many covariates that have been found to be associated with an infant's medical condition at birth. Research has shown, however, that none of these covariates accounts for a large portion of the variability in outcomes and, indeed, even combinations of these variables are not impressive in their predictive ability. The sole fact that a mother has had a previous low weight or preterm birth is as strong a predictor as any variable of the corresponding status of a current pregnancy. This observation suggests that perhaps a major part of the explanation for low weight and preterm births lies with the genomes of the infant or its parents and with the maternal fetal environment, which are not yet monitored routinely. The traditional variables (e.g., mother's age or race) probably exhibit their limited correlations with these birth conditions because of their weak associations with underlying genetic and environmental factors.

3 Premature birth as an intermediate endpoint

Low weight and preterm births are often referred to as a leading cause of infant morbidity and mortality. In fact, in relation to premature births, this exact phrase appears in the opening sentence of the Background section for Research Objectives in a recent NIH RFA call for proposals (National Institutes of Health 2004, p. 2). This kind of statement is representative of conventional thinking that claims these kinds of births are a *cause* of physiological and neurological problems for infants. The word 'cause', however, is inaccurate. A low weight or preterm birth may indicate a poor prognosis for future health outcomes of the infant but is not its cause. These conditions, where they are problems, are only symptoms or outcomes of deeper root causes that remain to be discovered and understood. Many authors have stressed this distinction between birth weight as a symptom rather than a cause but the point has not always been heeded (e.g., Wilcos 2001).

It is also important to recognize that birth weight or gestational age is, in fact, only an *intermediate endpoint* and not the final endpoint of concern. Our emphasis on the word 'intermediate' here is intentional. Many studies have analyzed births as if the birth weight or gestational age were the final endpoint of interest. Yet, babies who share the same short gestational age or low birth weight do not experience the same adverse health outcomes when they become children and adults. The problems that do arise vary in their form and severity. In addition, the adverse medical conditions of infants (e.g., forms of mental retardation) that are often associated with low weight or preterm birth are not limited to these kinds of cases.

In many studies of low weight or preterm birth, the outcome measure is a simple classifier of whether the gestational age is short or not (e.g., under 37 weeks) or whether the birth weight is low or not (e.g., under 2,500 g). Methods such as logistic regression are then used with these binary outcomes as response variables. This binary classification is simple but for research purposes is relatively insensitive. In our models, birth weight is taken as a continuous response variable that adjusts in a smooth fashion to continuously varying causal effects. Thus, for example, birth weight is expected to vary smoothly with covariates such as maternal age or alcohol consumption. Of course, some covariates are categorical variables (e.g., mother's race or birth delivery method). The benefit of handling birth weight as a continuous outcome is improved sensitivity and informativeness of response measurement. Other researchers have used conventional regression models that treat birth weight as a continuous response variable (see, for example, Kharrazi et al. 2004) but none of these models has the conceptual unity offered by our modeling approach.

In addition to more refined measures of gestational outcomes, our conceptual framework is enriched in several other respects. We have already pointed out the need to view low weight or preterm birth as an intermediate outcome. The final endpoints of interest are the physiological and neurological effects on the child and its development. Figure 1 shows the situation with a simple schematic. The figure shows that root causes are distinct from measurements (such as weight) that are made at birth. The figure also anticipates that some causal forces may escape detection in traditional birth measurements and have an impact on final outcomes directly. It may happen that other measures available at birth (besides gestational age and birth weight) may allow these bypass effects to be monitored. The Apgar score is one such measure that contains auxiliary information that is predictive of final outcome.

Low birth weight and premature birth are not synonymous terms. Moreover, a low birth weight or a short gestational age does not automatically signal a problem birth. Reproductive epidemiologists have engaged in much discussion and analysis centered on these two kinds of outcomes (e.g., Wilcox 2001; Savitz et al. 2002). Some babies with low birth weights will be in perfect health. Short gestational ages may arise because of measurement errors. Interuterine growth retardation (IUGR) is a term that has been invented to describe babies who are light in weight for their gestational age (e.g., the 10th percentile weight within each gestational age class). Still, even this kind of adjusted measure

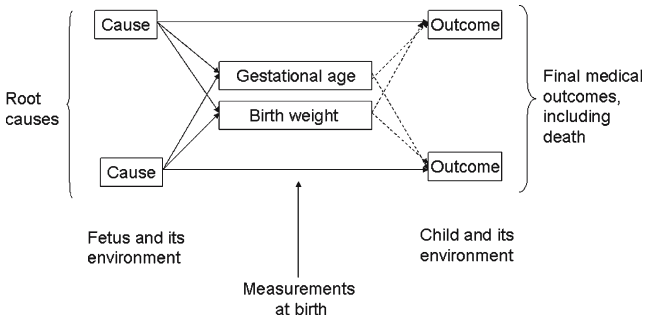

Fig. 1 Schematic showing gestational age and birth weight as intermediate endpoints positioned between root causes and final outcomes of prenatal and postnatal development. *Solid arrows* denote causal forces. *Dashed arrows* denote statistical associations. Neonatal death is one potential final outcome

does not quite capture the concept of an under-developed or premature newborn. The models we develop next will show that the modeling must be more sophisticated.

4 Modeling fetal and postnatal development

In our research, we model gestational age and birth weight as joint outcomes associated with a *fetal development process* $\{\mathbf{D}(t)\}$, where t denotes the time since conception and $\mathbf{D}$ is a multidimensional vector of physiological and other measures or features that characterize fetal development and the fetal environment. Some of the components of $\mathbf{D}$ will be observable directly or indirectly (e.g., fetal weight, fetal pulse), while others will be latent (unobservable) or unknown (e.g., brain development). We will assume that birth occurs when the fetal development process first reaches a birthing boundary $\mathbf{B}$ in the multidimensional space. Hence, the gestational age of the newborn T is a first hitting time or FHT, which is defined formally as follows:

$$T = \min\{t : \mathbf{D}(t) \in \mathbf{B}\} \tag{1}$$

The idea that a birth occurs when the fetal development process 'hits a boundary' is a mathematical abstraction that corresponds to the real-life circumstances that trigger birth. In this paper, we are interested in plausible but parsimonious models of the process $\{\mathbf{D}(t)\}$, the birthing boundary $\mathbf{B}$, and the dependence of both of these entities on covariates available from a conventional health statistics database. The purpose of the mathematical model is to allow all of the elements of the birthing process to be related in a logical fashion and to provide an analytical framework for studying the effects of explanatory variables, as will be described in more detail later. The point where the fetal development process strikes the birthing boundary defines the physiological and neurological characteristics of the newborn. In effect, the boundary

has different regions that represent different medical states of the newborn. For example, different Apgar scores would map probabilistically into different regions of the birthing boundary.

Fetal weight is one of the physiological measures in vector **D**, say component D_1, and therefore undergoes evolution after conception according to a process we call the *fetal weight process*. If $D_1(t)$ denotes the fetal weight at time t after conception and T is the gestational age at birth then $W = D_1(T)$ represents the birth weight. Figure 2 illustrates the situation. The schematic in the figure shows a two-dimensional fetal development process, with weight forming one dimension and some other unspecified physiological variable forming the other. Observe how birth is triggered when the process hits the boundary at time T after conception. That FHT event determines the stopping time of all components of the fetal development process, including the fetal weight. The physiological measures at birth are thus determined by the fetal development trajectory and its FHT.

It was stressed earlier that the condition at birth is an intermediate endpoint. Figure 2 reflects this viewpoint by showing how the development continues after birth as a postnatal development process for the child, essentially an analytical extension with a change in name. Of course, some components of the development process change as the physiology and environment of the fetus are not the same as those of the child. The continuing development of the child will be with or without medical complications, depending on individual circumstances.

The prenatal and postnatal development process will follow different sample paths in each individual case because of inherent variability. More importantly, the parameters of the process will vary systematically with characteristics of the fetus and child, and their prenatal and postnatal environments, respectively. This variation will be explained in part by the covariates in our data set. Also, the shape and position of the boundary **B** will also vary with these same cova-

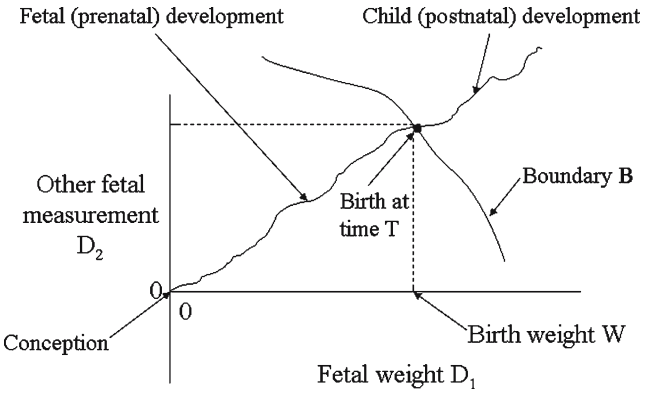

Fig. 2 The development process of the fetus and child in the prenatal and postnatal stages. Birth is represented as the first hitting time of a boundary by the development process

riates. This research studies the dependence of the model parameters on the explanatory variables during the prenatal stage.

5 Mathematical models

Figure 2 provides a general conceptual framework for the birthing process but is not specific enough to allow analysis and interpretation of data. In this section, we introduce model specifications that support a statistical investigation. At such a high level of abstraction, it is uncertain which model specifications are correct. Thus, we provide a start to the investigation by considering two parsimonious but realistic models.

5.1 Mixture model

We begin with the two-dimensional stochastic process illustrated in Fig. 2 where $\{\mathbf{D}(t)\} = \{D_1(t), D_2(t)\}$. Here $D_1(t)$ denotes the fetal weight at fetal age t and $D_2(t)$ denotes some other (unspecified) latent development measure at the same age. We shall refer to D_2 simply as the *latent fetal development measure*. We set both measures to zero at conception so $\mathbf{D}(0) = (0, 0)$. It is realistic to assume that the weight component $D_1(t)$ of the fetal development process has a continuous and monotonically increasing sample path with respect to gestational age t. Using this assumption, we will simplify the two-dimensional process to a one-dimensional subordinated process by defining a new process $\{C(u)\}$ as follows:

$$\{D_2(t)\} = \{C(D_1(t))\} = \{C(u)\} \tag{2}$$

where $u = D_1(t)$ denotes the weight of the unborn fetus at fetal age t. The fetal weight process $\{D_1(t)\}$ is a directing process here and the new stochastic process $\{C(u)\}$ is the parent process. Observe that $\{C(u)\}$ is the latent fetal development measure defined as a function of the fetal weight u. The sample path shown in Fig. 2 is, in fact, a representation of $C(u)$ where the horizontal scale is fetal weight u.

To give a specific form to the latent fetal development process $\{C(u)\}$, we assume it is a Wiener process with mean parameter $\mu > 0$ and unit variance. We specify a unit variance because $\{C(u)\}$ is a latent process and, hence, can be given an arbitrary measurement unit. We choose the Wiener process because it is mathematically tractable and flexible. It also exhibits the random fluctuation and drift that is characteristic of fetal development. For the birthing boundary $\mathbf{B}$, we assume that it corresponds to a fixed level $b > 0$ of fetal development in the $C(u)$ dimension. In this model, birth occurs at birth weight W where $C(W) = b$ for the first time. Thus, birth weight W is a first hitting time or FHT for a Wiener process. This FHT is known to have an inverse Gaussian distribution. We give its functional form later. Figure 3 shows a graphical representation of the FHT setup in this case.

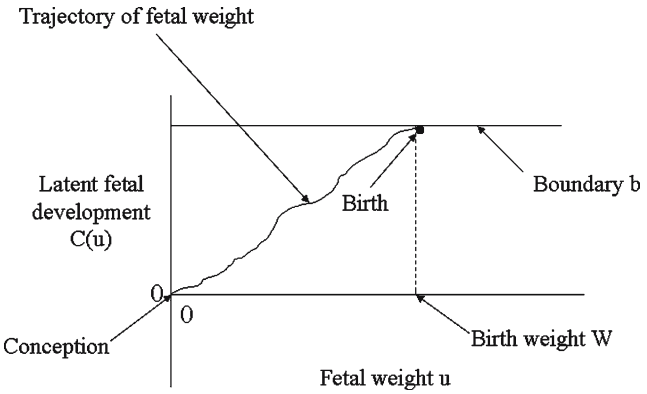

Fig. 3 The fetal development process $C(u)$ is a Wiener process defined in terms of fetal weight u. Birth is triggered when the sample path first hits a fixed boundary b. The birth weight W is defined by this event

We need to enrich our model somewhat to allow for the possibility that, while most fetuses will undergo *normal* development, a few are at risk of undergoing *abnormal* development. Recognizing this reality, we now consider the somewhat abstract but important idea that normal births (the vast majority) result when fetal development follows a *pristine natural course of evolution* from conception to birth. We assume that this natural course corresponds mathematically to the sample path of a stable stochastic process and, more specifically, to the Wiener FHT model that we have just described. In reality, of course, medical and other intentional interventions may interfere with this natural course for a birth and bring about an altered trajectory for process $\{\mathbf{D}(t)\}$ or an altered birthing boundary $\mathbf{B}$, which will imply that a simple model does not exactly fit real data. And, of course, we cannot be sure that our Wiener model is the right model (although evidence provided later is reassuring). In addition to normal births, there will be an array of abnormal births that arise from aberrant fetal development processes or birthing boundaries. These are births that do not occur under pristine natural conditions. The appropriate statistical model for abnormal births is likely to be more complicated than for normal births because of the variety of aberrations that occur. Yet, we will use the same kind of simple FHT model for abnormal births but assume that a different set of parameter values apply. We capture the combined influence of normal and abnormal births in the simplest possible way. Specifically, we postulate a population model for births that is a *birth weight mixture model*. Before we describe this mixture model, we wish to caution the reader that our use of the descriptors 'normal' and 'abnormal' is not intended to impart strong connotations to the words. The reader might loosely associate the words with 'healthy' and 'unhealthy', or 'natural' and 'unnatural'. The intention is to convey the idea that the birthing process has departed from the path dictated by healthy human biology.

Now, turning to the mixture model, we let N be an indicator variable for a normal birth so $N = 1$ if the birth is normal and $N = 0$ if it is not. We assume that every conception has a probability p_1 of leading to a normal birth and,

hence, probability $p_0 = 1 - p_1$ of leading to an abnormal birth. Subscripts 1 and 0 denote normal and abnormal, respectively. We further assume that whichever outcome holds, the fetal development path will follow a Wiener process and eventually hit a birthing barrier, producing the birth weight W. Each of these processes, however, will have different mean and boundary parameters, say, parameters μ_1 and b_1 for a normal case and μ_0 and b_0 for an abnormal case. Thus, if $F_1(w)$ and $F_0(w)$ denote the cumulative distribution functions (c.d.f.s) of the birth weight W for normal and abnormal cases, respectively, then the c.d.f. of birth weight $F(w)$ for the population of births is defined by the following mixture:

$$F(w) = p_1 F_1(w) + p_0 F_0(w). \tag{3}$$

Letting $f_1(w)$, $f_0(w)$ and $f(w)$ denote the corresponding probability density functions (p.d.f.s) of birth weight, then we have by differentiation of (3):

$$f(w) = p_1 f_1(w) + p_0 f_0(w). \tag{4}$$

The preceding notation has suppressed the dependence of the functions on the process parameters. Moreover, as we show later in applying threshold regression techniques, the parameters can be made to depend on covariates through regression link functions.

5.2 Competing risks model

We also propose a second model, which is conceptually distinct from the mixture model, but will tend to imitate birth weight data in a similar way. We refer to this alternate model as the *birth weight competing risks model*. In this model, we view the fetal development process as three-dimensional, as follows $\{\mathbf{D}(t)\} = \{D_1(t), D_2(t), D_3(t)\}$. We again let $u = D_1(t)$ denote the fetal weight u at fetal age t and redefine the fetal development process as a two-dimensional subordinated process as follows:

$$\{D_2(t), D_3(t)\} = \{C_1(D_1(t)), C_0(D_1(t))\} = \{C_1(u), C_0(u)\} \tag{5}$$

The components C_1 and C_0 of this subordinated process are, as before, latent fetal development measures that are expressed as functions of fetal weight. In the mixture model, a gamble at conception, with probabilities p_1 and $p_0 = 1 - p_1$, determines whether the fetus follows a normal or abnormal development. In contrast, in the competing risk model, we assume that the fetus develops continuously through time t in a two-dimensional space, undergoing stochastic movements in one dimension $\{C_1(t)\}$ that represents normal development and, simultaneously, undergoing stochastic movements in a second dimension $\{C_0(t)\}$ that represents abnormal development. As before, we associate subscripts 1 and 0 with normal and abnormal. We assume that $\{C_1(u), C_0(u)\}$ is a bivariate

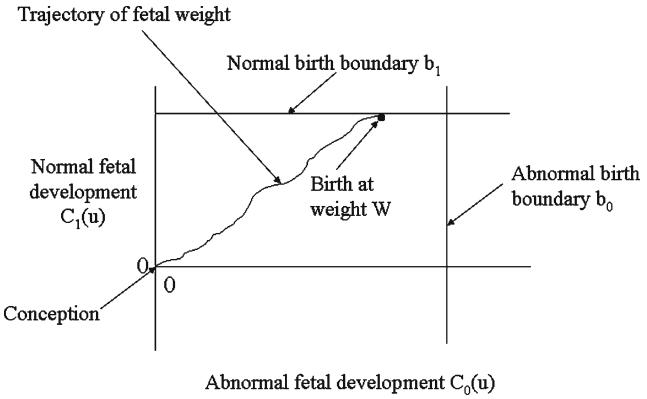

Fig. 4 The fetal development process $\{C_1(u), C_0(u)\}$ is a bivariate Wiener process defined in terms of fetal weight u. Birth is triggered when the sample path first hits one of the fixed boundaries b_1 or b_0 in the two dimensions. Whether the birth is normal or abnormal is determined by which boundary is encountered first. The birth weight W is defined by this event

Wiener diffusion process with mean parameters μ_1 and μ_0, respectively, and unit variances. The mean parameters μ_1 and μ_0 are not restricted in sign. We also assume that the respective boundaries are fixed levels $b_1 > 0$ and $b_0 > 0$ in the two dimensions. The two components C_1 and C_0 of the bivariate process are assumed to be independent or uncorrelated. Figure 4 shows a graphical representation of the FHT setup in this case.

Birth weight W for this bivariate process is defined as the smaller of the first hitting times of the two birthing boundaries, as follows:

$$W = \min\{W_1, W_0\} \qquad \text{where } W_N = \min\{u : C_N(u) \geq b_N\} \quad \text{for } N = 1, 0 \tag{6}$$

Thus, the birth weight W and condition of the newborn ($N = 0, 1$) is determined by the stochastic competition of these two development dimensions. Since the components of the bivariate process $\{C_1(u), C_0(u)\}$ are taken as independent, the first hitting times W_1 and W_0 in (6) are independent. Thus, if the c.d.f. of W_N is given by $F_N(u)$, for $N = 1, 0$, then the c.d.f $F(w)$ of W is given by

$$F(w) = 1 - [1 - F_1(w)][1 - F_0(w)]. \tag{7}$$

If the p.d.f.s of W_N and W are denoted by $f_N(w)$ and $f(w)$, respectively, then, by differentiating (7), we have that the p.d.f. of birth weight is

$$f(w) = f_1(w)[1 - F_0(w)] + f_0(w)[1 - F_1(w)]. \tag{8}$$

The preceding notation has suppressed the dependence of the functions on the process parameters. The formula for (8) has a similar structure to the corresponding formula for the mixture model given in (4), in that it shows a weighted sum of p.d.f.s $f_1(w)$ and $f_0(w)$. We study this correspondence more closely later

and discover that the two models are indeed quite close. As in the mixture model, the parameters can be made to depend on covariates through regression link functions.

5.3 Distribution formulas

We now present some formulas for the models that are required for later development. References for these formulas include, for example, Cox and Miller (1965) and Chhikara and Folks (1989).

Let the mean and variance parameters of a Wiener process be δ and ν, respectively. Assume a fixed boundary is located at $a > 0$. The first hitting time X has an inverse Gaussian distribution with the following p.d.f.:

$$h(x|\delta,\nu,a) = \sqrt{\frac{a^2}{2\pi\nu x^3}}\exp\left[-\frac{(\delta x-a)^2}{2\nu x}\right], \quad \text{for } -\infty<\delta<\infty \text{ and } \nu>0. \tag{9}$$

If $\delta < 0$ then the first hitting time is not certain to occur and the p.d.f. in (9) is improper. Specifically, in this case, $P(X=\infty) = 1-\exp(2a\delta/\nu)$. Conditioning on the first hitting time being finite, random variable X has the same p.d.f. as in (9) with δ replaced by $|\delta|$.

The cumulative distribution function (c.d.f.) corresponding to (9) is

$$H(x|\delta,\nu,a) = \Phi\left[\frac{(\delta x-a)}{\sqrt{\nu x}}\right] + \exp(2a\delta/\nu)\Phi\left[-\frac{(\delta x+a)}{\sqrt{\nu x}}\right], \tag{10}$$

where $\Phi(\cdot)$ is the c.d.f. of the standard normal distribution, which is a widely programmed mathematical function. For later reference, we mention that the mean and coefficient of variation (CV) of an inverse Gaussian variate X are

$$E(X) = a/\delta, \quad \mathrm{CV}(X) = \sqrt{\nu/(a\delta)} \quad \text{if } \delta>0. \tag{11}$$

One interesting property of an inverse Gaussian random variable X that we will use later relates to the following transformation of X.

$$Z = \frac{\delta X - a}{\sqrt{\nu X}} \tag{12}$$

Variable Z is a form of standardized inverse Gaussian variable and has the remarkable property that, for $\delta > 0$, the squared value Z^2 follows a χ^2_1 distribution, exactly like the square of a standard normal variable.

The components of the mixture model (4) and of the competing risks model (8) can be matched to the preceding formulas. For the normal birth component of the models, for instance, we equate w with x, μ_1 with δ, b_1 with a

and set $\nu = 1$ in the preceding formulas. Thus, $f_1(w) = h(w|\mu_1, 1, b_1)$ and $F_1(w) = H(w|\mu_1, 1, b_1)$. Similar formulas allow us to compute $f_0(w)$ and $F_0(w)$.

5.4 Comments on the models

The mixture and competing risks models have no built-in criterion that defines a low birth weight. We have given a mathematical structure to the models that accommodates two component distributions for birth weight but otherwise allow each model to be *self-calibrating* with respect to the discrimination between abnormal and normal birth weights. As will be seen later, the fitted models yield a plausible characterization of normal and abnormal birth weights; one that is quite close to the conventional definition of low birth weight. Moreover, the regression structure offers a refined distinction between abnormal and normal births. The models allow the definition of low birth weight, as characterized by the normal and abnormal birth weight distributions, to vary as a function of the covariates of the birthing circumstances.

Other researchers have recognized that the birth weight distribution might be modeled as a mixture of distributions. The model proposed by Wilcox and Russell (1983), for example, has generated considerable research and discussion. For a more recent discussion, see Wilcox (2001). This earlier research proposes a mixture model for the frequency distribution of birth weights which consists of a 'predominant distribution' that follows a normal p.d.f. and a 'residual distribution'. The predominant distribution represents the vast majority of births and, hence, corresponds to our normal birth weight distribution (distribution $f_1(w)$ in our mixture model). The residual distribution represents the remainder of births (mostly low birth weight cases) and, hence, corresponds to our abnormal birth weight distribution. The normal p.d.f. for the predominant distribution provides a remarkably close fit to the data and an adequate working model but is not totally satisfactory from a theoretical perspective because the normal p.d.f. admits negative outcomes. Our inverse Gaussian distribution can assume a shape that is very close to normal and does so for the birth weight component for normal births as our later results will show. Thus, in some respects, our mixture model is a mathematical refinement of this earlier mixture model. More importantly, perhaps, our model sets the mixture distribution within a mathematically coherent framework and also adds a regression structure to the analysis.

6 Statistical inference using threshold regression

6.1 Regression structure and link functions

Every birth has its own characteristics, some of which are captured by the huge array of characteristics monitored in health statistics records (such as maternal age, for example). We now introduce these characteristics as a covariate vector, say, vector $\mathbf{z} = (1, z_1, \ldots, z_k)$, where the leading 1 allows for a constant term

in regression models. We add a subscript to $\mathbf{z}$ and let vector $\mathbf{z}_i$ represent the covariate vector for birth i. In mixture model (4), these covariates may influence any of the five independent parameters of the model : $p_1, \mu_1, b_1, \mu_0, b_0$. Note that we have arbitrarily chosen parameter p_1 rather than $p_0 = 1 - p_1$ in this list. Letting symbol θ denote any one of these five parameters, we will choose a suitable mathematical link function to relate the parameter θ_i for birth i to a linear combination of its covariates $\mathbf{z}_i$, as follows:

$$g_\theta(\theta_i) = \mathbf{z}_i\boldsymbol{\beta},$$

where $\boldsymbol{\beta} = (\beta_0, \ldots, \beta_k)$'. Here the apostrophe denotes a vector transpose. The regression link function g_θ will be chosen for each parameter. Our choices map the parameter domains onto the whole real line. Thus, we choose $g_\theta(\theta)$ to be an identity function for μ_1 and μ_0, a natural logarithmic function $\ln(b)$ for b_1 and b_0, and a natural logit function $\ln[p_1/(1-p_1)]$ for p_1.

6.2 Model estimation

We estimate model parameters using the method of maximum likelihood. Letting $w_i, i = 1, \ldots, n$, be a sample of n independent birth weights and $\mathbf{z}_i, i = 1, \ldots, n$, their corresponding covariate vectors, the sample log-likelihood function is given by

$$\ln L(\boldsymbol{\theta}) = \sum_{i=1}^{n} \ln f(w_i|\mathbf{z}_i), \tag{13}$$

where $\boldsymbol{\theta}$ is the parameter vector for the particular model under consideration. The righthand side shows the sample data but, again, the parameter notation is suppressed. The maximum likelihood estimates are derived using a program written in version 7 of *Stata*. The optimization is carried out using a numerical gradient method, which works quite efficiently. Initial values for regression intercepts (one for each parameter) are chosen after a heuristic search for starting values that give a computable (finite) likelihood value. Finding such a feasible starting point is generally quite easy. All regression coefficients associated with covariates are initially set to zero (i.e., their values under the null hypothesis). Version 7 of the maximum likelihood routine in Stata also undertakes some internal scaling to improve the convergence speed of the numerical optimization routine. Likelihood functions for some mixture models have local maxima for small sample sizes. The large sample size here seems to eliminate this problem. We have introduced several well-separated starting points for fitting each of our regression models and all have converged to the same final estimates.

The two FHT models are fully identifiable when the components are mathematically distinct. The recognition of which component is associated with normal births and which with abnormal births is based on their relative positions

on the birth weight scale. Although model identification poses no problem, the regression parameters of these two FHT models have varying degrees of multicollinearity, which is a common condition in non-linear multivariate regression models. The appropriate parameterization remains a subject for future research.

6.3 Bayes analysis

The mixture and competing risks models have an implicit Bayesian interpretation. We consider the mixture model first and a birthing situation having covariate vector $\mathbf{z}$. We now incorporate the conditioning on the covariate vector explicitly in the notation.

Recall that N denotes an indicator variable for whether a birth is 'normal'. The mixture model (4) specifies $P(N = 1|\mathbf{z}) = p_1(\mathbf{z})$ as the prior probability that a birth with covariate vector $\mathbf{z}$ will be normal, without knowledge of the birth weight. It follows that $P(N = 0|\mathbf{z}) = p_0(\mathbf{z}) = 1 - P(N = 1|\mathbf{z}) = 1 - p_1(\mathbf{z})$ is the prior probability of an abnormal birth and the prior odds favoring an abnormal birth for a birth situation with covariate vector $\mathbf{z}$ are

$$\text{Odds}(\mathbf{z}) = \frac{P(N = 0|\mathbf{z})}{P(N = 1|\mathbf{z})} = \frac{p_0(\mathbf{z})}{p_1(\mathbf{z})} \tag{14}$$

If the newborn has weight $W = w$, then it follows that the posterior probability of an abnormal birth is given by

$$P(N = 0|w, \mathbf{z}) = \frac{p_0(\mathbf{z})f_0(w|\mathbf{z})}{f(w|\mathbf{z})} \tag{15}$$

and the posterior odds become

$$\text{Odds}(w, \mathbf{z}) = \frac{P(N = 0|w, \mathbf{z})}{P(N = 1|w, \mathbf{z})} = \frac{p_0(\mathbf{z})f_0(w|\mathbf{z})}{p_1(\mathbf{z})f_1(w|\mathbf{z})}. \tag{16}$$

A similar Bayesian interpretation can be given to the competing risks model (8). The respective prior and posterior probabilities for an abnormal birth are given by

$$P(N = 0|\mathbf{z}) = \int_0^\infty f_0(w|\mathbf{z})[1 - F_1(w|\mathbf{z})]dw, \tag{17}$$

$$P(N = 0|w, \mathbf{z}) = \frac{f_0(w|\mathbf{z})[1 - F_1(w|\mathbf{z})]}{f(w|\mathbf{z})}. \tag{18}$$

The integral defining the prior probability $P(N = 0|\mathbf{z})$ in (17) does not have a closed form and, hence, must be evaluated numerically for each case.

7 Case demonstration

7.1 Sample data set

We will examine the preceding models using the 4%-systematic sample of U.S. live births for 2002 mentioned earlier. The purpose of the case demonstration is not to provide a definitive analysis of U.S. birth data for 2002 but rather to explore the kinds of insights that might come from the mixture and competing risks models for birth weight using threshold regression. The paper uses a 4%-systematic sample because the intention is to provide only a case demonstration. The data are so extensive and involve so many health and medical technicalities that the requisite research and reporting will require several publications for other audiences. We have reserved 96% of the full data set for confirmatory analysis of more refined regression models that will be developed in further empirical research. The choice of 4%, rather than, say, 5% or some other fraction, is somewhat arbitrary. This 4% fragment includes 160,577 births, after eliminating 518 births with missing data—a data set that is more than large enough for our statistical purposes. Our systematic sample includes births numbered 1, 26, 51, etc. from the data file (i.e., every 25th birth starting from birth 1). The starting point was not a random choice.

7.2 Regression covariates and response variable

We had previously classified the variables in the NCHS data set according to their demographic and clinical characteristics. We now classify the variables according to their statistical characteristics as covariates. The classification is in keeping with the scheme presented in Fig. 2. The covariates in this sample data set fall into the following three categories: (1) baseline covariates, (2) interventions, and (3) covarying processes. *Baseline covariates* are characteristics that are fixed at or before the moment of conception and relate to maternal and paternal characteristics. *Interventions* describe actions that occur between conception and birth. *Covarying processes* are stochastic processes that operate in parallel with fetal development and are statistically associated with it (possibly because of direct causal influences). The following is a rough classification of the covariates in the NCHS database.

1. *Baseline covariates*: Mother's state and county of residence, numbers of previous live births (now living and dead), interval since last live birth, mother's age, live-birth order, mother's race, mother's Hispanic origin, mother's marital status, mother's educational attainment, father's age, father's race, father's Hispanic origin.
2. *Interventions*: Prenatal care, obstetric procedures, complications of labor and/or delivery, attendant at birth, method of delivery.
3. *Covarying processes*: Mother's weight gain during pregnancy, tobacco and alcohol use.

The classification of these variables is not always clearcut. The difficulties of classification are illustrated by covariates such as prenatal care that may be viewed as a covarying process in one form or as an intervention in another.

Several variables in the NCHS data set represent birth outcomes. *Outcomes* are characteristics of the newborn or the birth event that are revealed at the time of birth or shortly afterwards and are determined by the conception event and subsequent causal influences, both good and bad. The following is a list of outcome variables in the NCHS database.

Outcomes: State and county of birth, sex of newborn, date of birth, gestational duration, birth weight, Apgar score, abnormal conditions, congenital anomalies, multiple births.

As with covariates, the classification of variables as outcomes is not always clearcut. For example, the sex of the newborn might be taken to be a baseline covariate which can be known if desired. Moreover, practices that predetermine the sex of the newborn may disqualify this variable as an outcome measure.

We have selected the following covariates for the threshold regressions in this case demonstration. The first three are baseline covariates and the fourth is an intervention covariate.

1. *m_age*: Age of mother, in years (NCHS variable DMAGE).
2. *m_race*: Race of mother, coded 1 if white and 0 otherwise (one category of NCHS variable MRACE).
3. *prev_births*: Number of previous live births, whether now living or dead (corresponds to the sum of NCHS variables NLBLD and NLBND, representing numbers of live births, now living and now dead). The number excludes the current birth.
4. *vag_birth*: Method of delivery, coded 1 if vaginal without previous C-section and 0 otherwise (one category of NCHS variable DELMETH5).

The outcome of interest in this case demonstration, i.e., our response variable for threshold regression analysis, is birth weight, defined as follows.

b_weight: Birth weight, in kilograms. (NCHS variable DBIRWT, which is reported in grams)

Table 1 gives brief summary statistics for the covariates and response variable that we have selected for our regression analysis. These summary statistics assist in judging the potential magnitudes and ranges of regression effects. A look at the magnitudes of the Min and Max values for birth weight might suggest they are implausible. We have no basis, however, for concluding that any of the observations are in error. The minimum birth weight of 0.227 kg is exactly one-half pound. There is a slight concentration of eight cases at this weight but nearby values are also small. Moreover, eliminating the smallest values has only slight effects on the regression findings. Similarly, the largest value (7.002 kg) is well separated from the second largest (6.265 kg) but the mixture model has a slim right tail and such an extreme outcome is plausible. Again, omission of this extreme value from the regression analysis has a minute effect on the reported results.

Table 1 Summary statistics for the response variable and covariates used in the threshold regressions. All variables have 160,577 readings

Type	Variable		Statistics			
	Name	Units	Mean	SD	Min.	Max.
Response	*b_weight*	Kilograms	3.2984	0.60432	0.227	7.002
Covariates	*m_age*	Years	27.3618	6.18485	12	54
	prev_births	Births	1.0645	1.24266	0	17
	m_race	Indicator (0,1)	0.7908			
	vag_birth	Indicator (0,1)	0.7206			

8 Threshold regression results for case demonstration

Tables 2 and 3 show the threshold regression results for the mixture and competing risks models, respectively. The label *constant* represents the regression intercept in each case. The regression models in Tables 2 and 3 take no account of possible interactions (e.g., between mother's race and delivery method), curvilinear effects (e.g., for mother's age), or multiple births. The point estimates for regression parameters of the mixture and competing risks models shown in the tables are maximum likelihood estimates and the standard errors are asymptotic estimates derived from a numerical estimate of the negative inverse Hessian matrix for the likelihood surface at its maximum. The *P*-values assume that the parameter estimate vectors are multivariate normal.

Tables 2 and 3 include linear regression results for parameters $\ln(b_N)$ and μ_N for $N = 0, 1$. These results can be converted easily into regression results for the mean birth weight $E(W_N)$ and coefficient of variation of birth weight $\mathrm{CV}(W_N)$ for normal and abnormal births by using the relationships presented in (11). Since the dynamic ranges of the estimates for parameters $\ln(b_N)$ and μ_N over the data set are small, it follows that both $E(W_N)$ and $\mathrm{CV}(W_N)$ have nearly linear regression functions for the same covariates, for both normal and abnormal births.

8.1 Regression results for mixture model

Table 2 gives results for the mixture model. The comments that follow focus on statistically significant effects as shown by very small *P*-values.

1. The birthing barriers b_1 and b_0 for normal and abnormal births both tend to increase with mother's age (*m_age*), leading to larger birth weights (other factors being unchanged). For example, normal birth weights rise by about 0.14 percent for each year of age (0.001394 on the natural log-scale). There is also a hint that the probability of a normal birth p_1 declines slightly with age.
2. All parameters are positively associated with the mother's race being white (*m_race*), although this association is not firm for the mean parameter μ_0 of

Table 2 Threshold regression for a 4%-systematic sample of US birth weights in 2002, using the mixture model

Parameter	Variable	Estimate	SE	*P*-value
$\ln(b_1)$	*m_age*	0.001394	0.0003408	0.000
	m_race	0.087230	0.0055857	0.000
	prev_births	0.002328	0.0018432	0.207
	vag_birth	0.093775	0.0062312	0.000
	constant	2.385137	0.0121147	0.000
μ_1	*m_age*	−0.0024	0.001297	0.064
	m_race	0.135931	0.0207441	0.000
	prev_births	−0.006207	0.006932	0.371
	vag_birth	0.387638	0.0220058	0.000
	constant	3.526771	0.0448503	0.000
$\ln(b_0)$	*m_age*	0.004252	0.0011628	0.000
	m_race	0.210166	0.0178501	0.000
	prev_births	0.033335	0.0050999	0.000
	vag_birth	−0.140316	0.0163735	0.000
	constant	0.735839	0.0367198	0.000
μ_0	*m_age*	0.000002	0.0020418	0.999
	m_race	0.077311	0.0309055	0.012
	prev_births	0.018960	0.0088821	0.033
	vag_birth	−0.196094	0.0267596	0.000
	constant	1.288762	0.0649473	0.000
$\text{logit}(p_1)$	*m_age*	−0.007151	0.0028154	0.011
	m_race	0.217659	0.0375524	0.000
	prev_births	−0.128130	0.0133089	0.000
	vag_birth	1.275465	0.0357383	0.000
	constant	2.092567	0.0837875	0.000

abnormal births. The net effect of the simultaneous higher birthing barrier and mean parameter for births to white mothers is such that the mean birth weight is higher for both normal and abnormal births. The coefficient of variation (CV) of birth weights is smaller for babies born to white mothers for both types of births, implying that birth weights are relatively more concentrated for this race.

3. The birthing barrier for abnormal births b_0 increases with the number of previous live births for a mother (*prev_births*), which implies that the birth weights of abnormal births rise slightly with more previous births. On the other hand, the probability of a normal birth p_1 tends to fall with more previous births. Specifically, $\text{logit}(p_1)$ declines by an estimated −0.128 for each previous birth (other factors being unchanged).
4. The indicator variable for vaginal delivery without previous C-section (*vag_birth*) is associated positively with both parameters for normal births and negatively with both parameters for abnormal births. To interpret one regression coefficient for *vag_birth*, the regression coefficient of 0.093775 for $\ln(b_1)$, for example, indicates that the birthing boundary is about 10% higher, on average, with vaginal delivery (other factors being unchanged). Vaginal delivery is also strongly associated in a positive direction with the probability of normal birth p_1. Decreases in both the mean and CV of

Table 3 Threshold regression for a 4%-systematic sample of US birth weights in 2002, using the competing risks model

Parameter	Variable	Estimate	SE	*P*-value
$\ln(b_1)$	*m_age*	0.001748	0.0003562	0.000
	m_race	0.087879	0.0057006	0.000
	prev_births	0.005355	0.0019248	0.005
	vag_birth	0.072704	0.0062237	0.000
	constant	2.419734	0.0122377	0.000
μ_1	*m_age*	−0.001513	0.0013483	0.262
	m_race	0.139636	0.0212256	0.000
	prev_births	−0.003261	0.0071978	0.651
	vag_birth	0.361791	0.0219853	0.000
	constant	3.60822	0.0453646	0.000
$\ln(b_0)$	*m_age*	0.004558	0.0009199	0.000
	m_race	0.223207	0.0136575	0.000
	prev_births	0.042902	0.0041954	0.000
	vag_birth	−0.003503	0.0126968	0.783
	constant	0.583726	0.0284521	0.000
μ_0	*m_age*	0.007303	0.0016109	0.000
	m_race	0.176233	0.0216575	0.000
	prev_births	0.096471	0.0074348	0.000
	vag_birth	−0.389698	0.0209808	0.000
	constant	−0.443521	0.0479385	0.000

normal birth weights are associated with vaginal delivery. The opposite happens with abnormal births, with increases in both measures.

5. By averaging the estimates of $p_0(\mathbf{z}) = 1 - p_1(\mathbf{z})$ for all cases, we obtain an estimate of the population proportion of abnormal births. This average is 0.065 or 6.5% for the mixture model. The individual estimates of $p_0(\mathbf{z})$ range from 3.0% to 49.3% as the covariates vary over the sample cases. As an interesting comparison, we note that 7.8% of all birth weights are under 2,500 g, the usual cutoff for low birth weights.
6. The Apgar score is a routine evaluation of the newborn on five clinical parameters, namely, heart rate (pulse), breathing (rate and effort), activity and muscle tone, grimace response (reflex irritability), and appearance (skin coloration). Each parameter is assigned a value from 0 to 2, with a total score ranging from 0 to 10. A score of 9 or 10 indicates excellent health; 7 or more, good health. The Apgar score may be considered as the outcome of the newborn's first clinical examination. We consider the score that is recorded five minutes after birth. All states reported 5-min Apgar scores for births in 2002 except California and Texas. In total, Apgar scores were available for 77% of births in our sample. Table 4 shows the mean of the estimated probabilities $P(N = 0|w, \mathbf{z})$ for abnormal birth derived from our fitted regression functions for the mixture model, classified by several Apgar categories that are traditionally used to grade a birth. None of the parameters of the Apgar score relates directly to birth weight or to any of the covariates that have been used in our illustrative regression model. Thus, it is an independent

Table 4 Relation of the Apgar score to the estimated probability of abnormal birth, averaged for all cases in each category. Estimates are derived from the fitted mixture model

Apgar score	Missing	0–3	4–6	7–8	9–10
Number of cases	36,585	513	1185	10,417	111,877
Mean probability $P(N = 0\|w, \mathbf{z})$	0.060	0.729	0.390	0.192	0.048

outcome variable. As expected, we see that the estimated mean probability of abnormal birth descends steeply as we move to higher Apgar categories. The table confirms a strong association of clinical condition with the model's estimated probability of abnormal birth. The association is not strong enough to make the Apgar score a strong predictor of whether a birth is abnormal of not but it does show that the mixture model has clinical validity and that the Apgar score might be a useful supplemental outcome variable. We comment a little more on this issue in the later section *Birth Weight z-score*.

8.2 Regression results for competing risk model

Table 3 gives results for the corresponding competing risks model. When appropriately interpreted, the results are not much different than those for the mixture model. The following remarks focus on a comparison and contrast of the regression results in the two tables.

1. The estimated distributions of normal birth weight are nearly the same as that of the mixture model across the range of covariate values. This fact is shown by the almost identical estimated regression functions for parameters $\ln(b_1)$ and μ_1. This result is reassuring because it implies that the two models capture this feature of the data set in a similar way.
2. The estimated distributions of abnormal birth weight in the two models are not as close as they are for normal weights, but the comparison is somewhat complicated for the following reasons.
   (a) The mean parameter μ_0 for abnormal births is negative for about 85% of births. A negative value for μ_0 implies that the fetal development process tends to drift away from the abnormal birthing barrier, making abnormal birth a less competitive risk. In this situation, an abnormal birth has a positive probability of never occurring. This drift away from the abnormal birthing barrier is a major determinant of the percentage of normal births because of the reduced risk of an abnormal outcome.
   (b) The mixture model has an extra parameter (parameter p_1) relative to the competing risks model. Looking at the tables, we see that the regression structures for parameter $\ln(b_0)$ are roughly similar for the mixture and competing risks models, except for the indicator variable *vag_birth* which has an insignificant regression coefficient in the competing risks model. The regression coefficients of the covariates for the mean parameter

μ_0 have matching signs to those of the mixture model but are larger in magnitude. The constant terms for μ_0 in the two models differ markedly. The reason for these shifting regression coefficients for the abnormal birth parameters is the fact that the two parameters $\ln(b_0)$ and μ_0 of the competing risks model must capture the effects of the three parameters $\ln(b_0)$, μ_0 and p_1 of the mixture model. As we will see later, however, the net effects of these shifts virtually wash out and yield almost identical overall fits to the birth weight distribution. In terms of the number of independent model parameters, the competing risks model is more parsimonious.

3. The effect of vaginal delivery deserves an extra comment for the competing risks model. As with the mixture model, the indicator variable for vaginal delivery without previous C-section (*vag_birth*) has a positive association with both parameters for normal births and a negative association with the mean parameter for abnormal births. The net effect is that the mean and CV of birth weights tend to be smaller for normal births. A substantial decrease in the incidence of abnormal birth weights is associated with vaginal delivery because the regression coefficient for μ_0 is large and negative.

8.3 Regression results for one scenario

Interpretation of some results in Tables 2 and 3 requires consideration of specific scenarios for baseline and intervention covariates. To demonstrate, we arbitrarily choose one scenario, namely, that of a 30-year-old non-white mother with no previous live birth and a C-section delivery for this birth. Thus, $m\_age = 30$, $m\_race = prev\_births = vag\_birth = 0$. These define covariate vector $\mathbf{z}$. In our data set, 175 births match this scenario. We have the following results under the two models for this scenario.

1. The prior probability $P(N = 1|\mathbf{z}) = p_1(\mathbf{z})$ is estimated to be 0.867 for the mixture model. Thus, an estimated 87% of births are from the normal component according to this model. Interestingly, the prior probability (17) gives almost the same probability for the competing risks model, namely, $P(N = 1|\mathbf{z}) = 0.845$.
2. The estimated probability densities $f_0(w|\mathbf{z})$ and $f_1(w|\mathbf{z})$ are approximately equal at birth weight $w = 2.45$ or 2,450 g under the mixture model. This corresponds closely to the conventional cutoff of 2,500 g for low weight births. For the competing risks model, the two densities are equal at 2,260 g, a little lower value than for the mixture model. Alternatively, we can also look at the posterior probabilities. The posterior probability $P(N = 1|w, \mathbf{z})$ in (15) is 50% when $w = 2.16$ or 2,160 g under the mixture model. The corresponding probability in (18) is 50% when $w = 2.27$ or 2,270 g under the competing risks model. The posterior probability suggests that the odds favor a declaration of normal birth when the birth weight is above 2,160 or 2,270 g, depending on the chosen model. Both of these cutoffs are lower than the conventional value of 2,500 g.

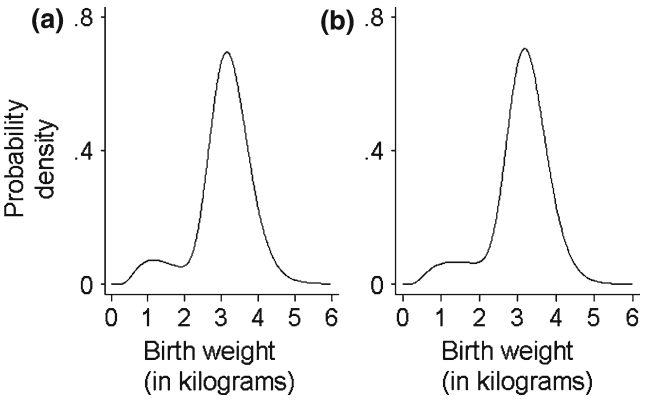

Fig. 5 The estimated probability density function for birth weight under (**a**) the mixture model and (**b**) the competing risks model, for 30-year-old non-white mothers with no previous live birth and a C-section birth ($m\_age = 30$, $m\_race = prev\_births = vag\_birth = 0$)

3. The estimated mean birth weights for normal births under the mixture and competing risks models for this scenario are 3,280 and 3,330 g, respectively. In the case of the competing risks model, the parameter is interpreted as the mean that would prevail without competition from abnormal births.
4. Panels (a) and (b) of Fig. 5 show the estimated probability density functions for birth weight under the mixture and competing risks models for this scenario, respectively. The two models have produced very similar estimates. They differ slightly at the junction of the abnormal and normal components.

9 Model checking

Three basic sources of error must be considered in model checking: (1) deviation from the true model caused by sampling error, (2) an incorrectly specified parametric model and (3) an incorrectly specified regression structure for the parametric model. Figure 6 illustrates these sources of error for one birthing scenario in our case demonstration. The figure is derived from the 2,836 births to 19-year-old white mothers who are having their first birth by vaginal delivery ($m\_age = 19$, $m\_race = 1$, $prev\_births = 0$, $vag\_birth = 1$). This subset of the sample happens to be the modal configuration of these covariates. The figure compares the empirical birth weight c.d.f. for this subset with two fitted c.d.f.s derived from the mixture model. In panel (a), the mixture model is fitted using the regression structure of the case demonstration, which is estimated from all 160,577 births in the sample and has the estimates displayed in Table 2. Small gaps are visible between the empirical and fitted c.d.f.s. The differences are attributable to all three sources of error. In panel (b), the mixture model is fitted using a saturated regression model. In this instance, estimation of the model parameters is carried out without regression covariates using only the 2,836 births in the scenario subset. (The parameter estimates are not presented

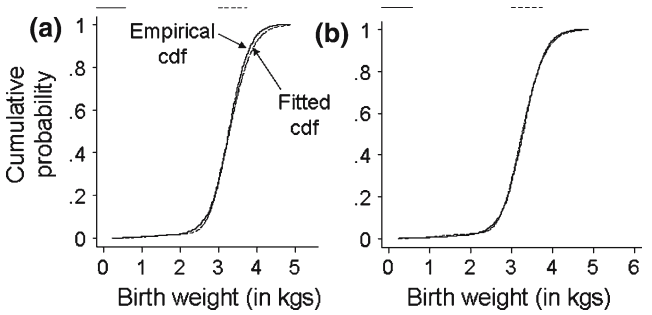

Fig. 6 A comparison of an empirical c.d.f. of the birth weight distribution with a fitted c.d.f. based on the mixture model for (**a**) the regression structure of the case demonstration and (**b**) a full-model estimate (a saturated regression model). The comparison is for 2,836 births of 19-year-old white mothers having their first birth by vaginal delivery ($m\_age = 19$, $m\_race = 1$, $prev\_births = 0$, $vag\_birth = 1$)

here.) The empirical and fitted c.d.f.s now coincide quite closely. The saturated regression model is equivalent to using the true regression structure and, thus, eliminates the third source of error. The figure is typical of what is found for other subsets (although others involve smaller numbers of births and, hence, larger sampling errors). Figure 7 compares the empirical and fitted survival probabilities in a probability plot that highlights the differences found in panel (b) of Fig. 6. The differences in the c.d.f.s range from −0.037 to 0.016 and have some regularity. For example, the differences have a slight wave in the lower region of the probability plot. The wave suggests that the fitted model predicts more very low birth weights than appear in the data set. As the data set considers only live births, the discrepancy may very well reflect the absence in the data of weights for still births. We discuss this issue further in the last section.

The model checking results suggest that the mixture model is reasonably good but needs some fine-tuning. The regression structure in the case demonstration, on the other hand, needs much further development in terms of the choice of covariates and/or the selection of their mathematical forms in the regression model. For example, we know that mother's age (covariate *m_age*) needs to appear in the model with a curvature effect. The model-checking results for the competing risks model are similar but not identical. More research is needed to determine which of these models is superior or to discover if some refined hybrid model is better than either one of them. For the present, however, the results in Figs. 6 and 7, and similar results for other scenarios in the data set, suggest that our modeling efforts are moving in the right direction.

Tables 2 and 3 provide point estimates, asymptotic standard errors and *P*-values for parameters of the mixture and competing risks models. To confirm the validity of these asymptotic results, we have also obtained inferences using a more robust procedure, namely, inferences from subsamples. Specifically, the regression models were re-run for 32 non-overlapping and equal subsamples of our 4%-systematic sample. By this construction, the 32 sets of regression estimates are identically distributed draws from the same underlying sampling

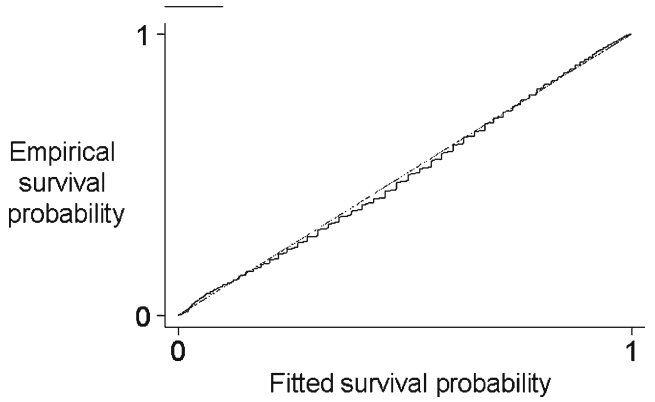

Fig. 7 A probability plot comparing the empirical and fitted survival probabilities for the saturated regression model shown in panel (b) of Figure 6. The comparison is for 2836 births of 19-year-old white mothers having their first birth by vaginal delivery ($m\_age = 19$, $m\_race = 1$, $prev\_births = 0$, $vag\_birth = 1$)

distribution. By calculating the means of the parameter estimates and testing whether they differ significantly from zero, we obtain a check on the reliability of our previous *P*-values. The number of subsamples was set at 32 because 32 evenly divides the sample into adequately large subsamples (about 5,000 observations each) while maintaining a reasonable number of degrees of freedom ($df = 32 - 1 = 31$) for the tests. These sample subsets provide robust estimates and inferences for model parameters that do not depend greatly on the underlying sampling error properties of the models. Table 5 gives the mean parameter estimates and *P*-values for each model. Comparisons with the full-sample maximum likelihood estimates and *P*-values in Tables 2 and 3 shows generally good agreement with respect to magnitudes of significance effects. The subsample *P*-values for parameter $\ln(b_0)$ for the mixture model are larger for several covariates which suggests that the parameter estimates are somewhat over-dispersed among the subsamples. Also, under the competing risks model, the level of significance of the effect of *m_age* on μ_0 shows some erosion under the subsample procedure. We note that this subsampling method gives robust estimates of variance that are much like those produced by so-called sandwich estimators.

10 Birth weight *z*-score

We noted earlier that Wilcox and Russell (1983) proposed a mixture model in which the 'predominant distribution' (our birth weight distribution for normal births) has a normal p.d.f.. They then point out the value of looking at a *z*-score, the number of standard deviations that a birth weight lies from the mean of the predominant distribution for a given reference group. The corresponding score

Table 5 Regression results for 32 subsets of the 4%-systematic sample of US birth weights in 2002 for the mixture and competing risks models. The table reports the mean estimates and *P*-values of tests for zero effects based on *t*-tests with 31 degrees of freedom

Parameter	Variable	Mixture model		Competing risks model	
		Mean estimate	*P*-value	Mean estimate	*P*-value
$\ln(b_1)$	*m_age*	0.001421	0.001	0.001655	0.000
	m_race	0.087581	0.000	0.086904	0.000
	prev_births	0.003235	0.213	0.006394	0.017
	vag_birth	0.090116	0.000	0.070776	0.000
	constant	2.391919	0.000	2.428129	0.000
μ_1	*m_age*	−0.002307	0.127	−0.001825	0.221
	m_race	0.137048	0.000	0.135699	0.000
	prev_births	−0.002687	0.766	−0.000231	0.980
	vag_birth	0.375088	0.000	0.354598	0.000
	constant	3.552884	0.000	3.639945	0.000
$\ln(b_0)$	*m_age*	0.005448	0.028	0.004840	0.010
	m_race	0.213401	0.000	0.219821	0.000
	prev_births	0.034156	0.005	0.045276	0.000
	vag_birth	−0.112027	0.002	0.010872	0.624
	constant	0.714408	0.000	0.584972	0.000
μ_0	*m_age*	0.001603	0.506	0.007507	0.016
	m_race	0.080888	0.003	0.177970	0.000
	prev_births	0.028457	0.031	0.103805	0.000
	vag_birth	−0.166199	0.000	−0.366020	0.000
	constant	1.262559	0.000	−0.444823	0.000
$\text{logit}(p_1)$	*m_age*	−0.007892	0.012	–	–
	m_race	0.194776	0.000	–	–
	prev_births	−0.129026	0.000	–	–
	vag_birth	1.257793	0.000	–	–
	constant	2.120913	0.000	–	–

for our mixture model is given by (12), which takes the following form in our notation:

$$z = \frac{\mu_1 w - b_1}{\sqrt{w}} \tag{19}$$

This z-score is zero when the birth weight equals the mean for normal births and roughly behaves like a standard normal number, as does the Wilcox and Russell z-score. Wilcox (2001) points out that when infant mortality is plotted against the z-score scale for different conditions, one can make a valid comparison of infant mortality rates as a function of birth weight. The underlying reason is that the z-score standardizes the predominant distribution so it is invariant for comparison groups. Thus, observed differences in infant mortality must be attributable to what they call the 'residual distribution' (our birth weight distribution for abnormal births). We are not studying infant mortality here but if we plot the posterior probability of abnormal birth against our z-score, as defined in (19), we obtain a plot that is similar to what appears in Wilcox (2001). Figure 8 shows such a plot for 19-year-old white mothers

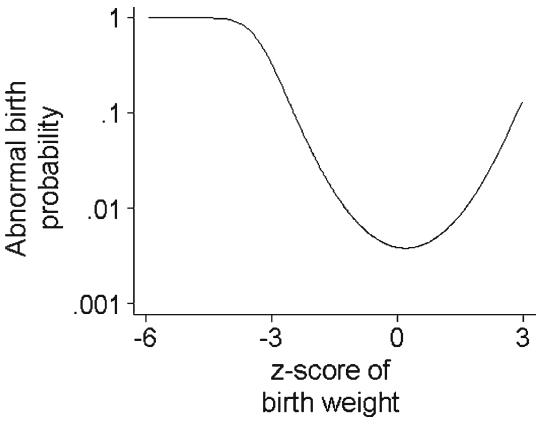

Fig. 8 Posterior probability of an abnormal birth plotted against the z-score of the birth weight based on the transform in (19). The vertical scale is logarithmic. The plot is for 19-year-old white mothers having their first birth by vaginal delivery ($m\_age = 19$, $m\_race = 1$, $prev\_births = 0$, $vag\_birth = 1$)

having their first birth by vaginal delivery (m_age=19, m_race=1, $prev\_births = 0$, vag_birth=1). This group was featured in our earlier discussions. The horizontal scale of the figure shows our z-score. The vertical scale is a logarithmic scale that shows the posterior probability of an abnormal birth, based on the fitted mixture model. This probability corresponds to $P(N = 0|w, \mathbf{z})$ in (15). A plausible feature that is evident from the figure is that the risk of an abnormal birth is present at every birth weight. The risk is small in the neighborhood of zero on the z-score scale but does not vanish. We also see a feature in the plot that was not noted earlier, namely, the tendency of the probability of an abnormal birth to rise when birth weights are well above the mean. Wilcox noted that infant mortality rises slightly at large birth weights. Our mixture model implies that large birth weights may arise from abnormal fetal development and birthing conditions. The fact that these abnormal cases do not tend to be life threatening for the newborn may explain why infant mortality does not rise as sharply in the upper weight range as our graph shows. In connection with Table 4 given earlier, we note that Apgar scores also show a tendency to drop at large birth weights, which adds clinical weight to the observation that the probability of an abnormal birth is higher at large birth weights.

11 Birth weight and gestational age

Gestational age, denoted earlier by the first hitting time T in Fig. 2, is another outcome measure that is frequently used as an indicator of premature birth. We have postponed consideration of this response measure to future research but have a few remarks to make at this point.

The NCHS variable for gestational age (GESTAT in their data base) is (a) computed using the birth date and the date of the last normal menses, (b) imputed from the date of the last normal menses, (c) estimated clinically, or (d) set to 'unknown' when the data are insufficient to impute the age or no valid clinical estimate is available. Gestational age is therefore subject to potential

measurement error that should be taken into account in analysis. The uncertain time of conception might be accommodated as a random variable T_0 so the state of the fetal development process at calendar time t is given by $\mathbf{D}(t - T_0)$. One advantage of birth weight as an outcome measure is that measurement error for date of conception does not affect it. The fetal weight is zero at conception and, hence, the origin of the birth weight scale is known. The measurement error in gestational age and the discrete recording of gestational age (in whole weeks) require special handling.

We recognize that the joint distribution of birth weight and gestational age represents a more refined outcome profile for births than either measure taken by itself. To give a sense of what would be required in a study of the joint outcome of birth weight W and gestational age T (see Fig. 2), one would need to consider, say, a mixture of two bivariate distributions defined for the joint outcome (W, T). This mixture would have a dominant component for normal births and a second component representing abnormal births. In birth weight studies, the birth weight variable is often adjusted for gestational age. In effect, this adjustment involves a study of the conditional outcome variable $W|T$. Inferences drawn from a study of $W|T$ are conceptually distinct from those drawn from the unconditional variable W. For example, the conditional birth weight $W|T$, adjusted for gestational age, is likely to depend on covariates differently than the unconditional birth weight W. It is the latter that is being studied in our case demonstration. Other outcome measures, such as the newborn's Apgar score, can be added to W and T to further refine the outcome profile. These topics requires further investigation.

12 Concluding remarks

The NCHS data set we have used includes only live births and, therefore, is truncated with respect to prenatal conditions that lead to still births or spontaneous abortions. In essence, these events are unaccounted for and ignored in our models and inferences. The events affect the abnormal birth component of the models and may explain some of the model discrepancies found in Fig. 7. Our models can be extended to include still births or spontaneous abortions if these data become available. The basic mathematical extension involves viewing conceptions as having competing outcomes in the form of being a live birth, still birth or spontaneous abortion. Our modeling also ignores the effects of multiple births. The data show clear evidence that the total weight of multiple births increases with multiplicity but at a declining rate. As a consequence, birth weight for each infant declines with multiplicity—see NCHS (2003), Table K, p.22. Our models can be extended to include a probabilistic component for multiple births. This extension is left to future research.

One observation that stands out clearly in our results is the striking differential effects on birth weights associated with vaginal and cesarean delivery methods. NCHS birth data show that the cesarean rate is rising. As noted in NCHS (2002, p. 15), "The escalation in the total cesarean rate is fueled by both

the rise in the primary cesarean rate and the steep decline in the rate of vaginal birth after cesarean (VBAC) delivery. Controversy continues to stimulate research and discussion on the risks, benefits, and long-term consequences of cesarean (medically indicated or elective) delivery and VBAC delivery...". See Lydon-Rochelle et al. (2001), Gregory et al. (2001) and Scott (2002).

Mixture models have been used to describe population distributions for other physiological features. For example, Harlow and Zeger (1991) proposed that the menstrual cycle of women is a mixture of a dominant normal pattern and an abnormal pattern. In a recent study, Guo et al. (2006) use a mixture distribution for menstrual cycles consisting of a normal distribution for 'standard cycles' and a shifted Weibull distribution for 'nonstandard cycles' A mixture of two inverse Gaussian distributions has been proposed in other settings. Balka (2005), for example, has proposed such a model for survival data with a cure rate.

The *International Journal of Epidemiology* devoted a recent issue to the topic of obesity. In an editorial, Lawlor and Chaturvedi (2006) speak of the critical influence of the perinatal period on obesity in later life. They observe that the mechanisms of association between maternal weight and weight gain during pregnancy and obesity in offspring are becoming clearer. In the same issue, Chen et al. (2006) show a strong association between maternal smoking during pregnancy and early childhood obesity. The model we propose here provides an ideal structure for the examination of the link between maternal weight and smoking habits and birth weight. Extension of the model to consider health outcomes in early childhood (such as obesity) would add greatly to its usefulness. The connection of birth weight and other birth outcomes to childhood obesity requires much further study.

Our aim in this research report is to use modeling and analysis to understand better the statistical nature of low weight births. Much research remains to be done and the NCHS data base is an exceptionally good resource for this work. Our research postulates a model for the development trajectory of a fetus and its consequent birth weight. The model distinguishes between normal and abnormal birth trajectories as reflected in birth weights. We then use threshold regression to examine the correlation of model parameters with various covariates that are recorded routinely in health statistic databases for births. Some of the practical advantages of our modeling approach may be summarized as follows.

1. The approach offers a rich and unified conceptual framework for future research on fetal development, birth and, even, early childhood development. For instance, the model can be used when measurements of fetal growth become available (e.g., measurements derived from ultrasound images of the fetus). The model may also be extended to correlate birth outcomes with early childhood development (e.g., the connection to obesity).
2. There is great value in knowing which covariates are associated with the parameters $\ln(b_0)$ and μ_0 of the abnormal component of the model. Likewise, in the case of the mixture model, the influence of covariates on parameter p_0, representing the proportion of abnormal births, is of interest.

3. The new z-score defined in (19) is an important practical tool for examining and comparing abnormal births in different reference populations.
4. The model does not provide a forecast of whether any birth is normal or abnormal. But it does provide a probability for this event, conditional on birth weight w and the given covariate values $\mathbf{z}$ for the case—see expressions (15) and (18).
5. As the model and regression structure are completely general, the proposed model can accommodate in-depth investigations of associations between birth weight and covariates that go well beyond the case demonstration reported here. The extension to other outcome measures (such as gestational age) is natural and conceptually straightforward.

Acknowledgements This research is supported in part by a research grant from the Natural Sciences and Engineering Research Council of Canada (Whitmore). The authors thank the National Center for Health Statistics of the United States for making their birth data available for researchers in a CD format, together with ample supporting documentation. The authors also thank four reviewers for excellent suggestions that have greatly improved the manuscript. Any errors or deficiencies that remain, however, are solely the responsibility of the authors.

References

Balka J (2005) Wiener process cure rate models, Ph.D. thesis, University of Guelph, Guelph, Canada

Chen A, Pennell ML, Klebanoff MA, Rogan WJ, Longnecker MP (2006) Maternal smoking during pregnancy in relation to child overweight: follow-up to age 8 years. Int J Epidemiol 35:121–130

Chhikara RS, Folks JL (1989) The inverse Gaussian distribution: theory, methodology and applications. Marcel Dekker, New York

Cox DR, Miller HD (1965) The theory of stochastic processes. Chapman & Hall London

Gregory KD, Korst LM, Platt LD (2001) Variation in elective primary cesarean delivery by patient and hospital factors. Am J Obstet Gynecol 184(7):1521–1534

Guo Y, Manatunga AK, Chen S, Marcus M (2006) Modeling menstrual cycle length using a mixture distribution. Biostatistics 7:100–114

Harlow SD, Zeger SL (1991) An application of longitudinal methods to the analysis of menstrual diary data. J Clin Epidemiol 44:1015–1025

Kharrazi M, DeLorenze GN, Kaufman FL, Eskenazi B, Bernert JT Jr, Graham S, et al. (2004) Environmental tobacco smoke and pregnancy outcome. Epidemiology 15:660–670

Lawlor DA, Chaturvedi N (2006) Treatment and prevention of obesity—are there critical periods of intervention, Editorial. Int J Epidemiol 35:3–9

Lee M-LT, DeGruttola V, Schoenfeld D (2000) A model for markers and latent health status. J Roy Stat Soc, Ser B 62:747–762

Lee M-LT, Garshick E, Whitmore GA, Laden F, Hart J (2004) Assessing lung cancer risk to rail workers using a first hitting time regression model. Environmetrics 15:1–12

Lee M-LT, Whitmore GA (2004) First hitting time models for lifetime data. Handbook of statistics, Vol 23. Elsevier, pp 537–543

Lee M-LT, Whitmore GA (forthcoming) Threshold regression for survival analysis: modeling event times by a stochastic process, *Stat Sci*

Lydon-Rochelle M, Holt VL, Easterling TR, Martin DP (2001) First-birth cesarean and placental abruption or previa at second birth. Obstet Gynecol 97(5):Part 1. 765–769

National Institutes of Health (2004) Genomic and proteomic network for premature birth research, release date, June 18, 2004, RFA-HD-04-002, Department of Health and Human Services (DHHS)

National Center for Health Statistics (1980) Factors associated with low birth weight, United States, 1976, DHEW Publication No. (PHS) 80–1915

National Center for Health Statistics (2003) Births: final data for 2002, National Vital Statistics Reports 52(10), December 17, 2003
Savitz DA, Hertz-Picciotto I, Poole C, Olshan AF (2002) Epidemiologic measures of the course and outcome of pregnancy. Epidemiologic Reviews 24:91–101
Scott JR (2002) Putting elective cesarean into perspective. Editorial. Obstet Gynecol 99(6):967–968
Whitmore GA (1983) A regression method for censored inverse-Gaussian data. Can J Stat 11: 305–315
Wilcox AJ, Russell IT (1983) Birthweight and perinatal mortality. I. On the frequency distribution of birthweight. Int J Epidemiol 12:314–318
Wilcox AJ (2001) On the importance—and the unimportance—of birthweight (with commentary). Int J Epidemiol 30:1233–1241