Mixed Logistic Regressions with Covariate Density Defined Components: Applications to Birth Outcomes

By

Timothy B. Gage 1,2,4

Michael J. Bauer 3

Fu Fang 3

and

Howard Stratton 3

1 Department of Anthropology
2 Department of Epidemiology
3 Department of Biometry
University at Albany
Albany, NY 12222

4 Department of Genetics
Southwest Foundation for Biomedical Research
P.O. Box 760549
San Antonio, TX 78245

Send correspondence to:

Timothy B. Gage
SS 259
University at Albany – SUNY
Albany, NY 12222

Office 518-442-4704
Fax   518-442-5710

Abstract

The statistical properties of covariate density defined, (CDD) mixtures of logistic regressions as a method of controlling for heterogeneity in infant mortality are explored. Unlike finite mixtures of logistic regressions, the CDD approach is usually identified and is probably generalizable to most regression like procedures. CDD mixtures use the marginal density of a covariate (birth weight in this case) to assign probabalistic (latent) group membership to separate logistic probabilities. The procedure appears to be unbiased, and consistent. A procedure for estimating power is presented. The method identifies significant heterogeneity, which influences birth weight specific infant mortality, and is consistent across populations. This heterogeneity is the proximate cause of the "pediatric paradox", i.e. the finding that low birth weight African American infants have lower infant mortality then European American infants. All of the "paradox" occurs in one subpopulation. Applications with additional covariates could identify the ultimate causes of this heterogeneity.

Introduction

Finite mixture models are increasingly used as a cluster analysis technique with the results interpreted as heterogeneous subpopulations within a larger population, e.g. the effects of sex or age etc (McLachlan and Peel 2000). Such interpretations are clearly relevant when the existence of subpopulations and the number of subpopulations are known *a priori*, e.g. sex in a heterosexual population. However, finite mixture models have also been used to infer the existence of subpopulations, which are not known *a priori*. For example Pearson (1894) in the earliest application of a two-component finite Gaussian mixture model argued that crabs from the Bay of Naples might consist of two subspecies. This interpretation of finite mixture models, particularly where the existence of subpopulations are not known *a priori*, suggests that finite mixture models might provide a useful method of statistically controlling for otherwise unobserved or unobservable heterogeneity in the context of various types of regression. However, finite mixtures of general linear models are not typically specified using a covariate density defined finite mixture submodel of this type.

Finite mixtures of general linear models are traditionally specified assuming that the subpopulation structure is unknown and that support for the subpopulation structure is obtained from the dependent variable alone (McLachlan and Peel 2000; Wang 1994). Typically these models are defined as the sum of several component-specific regressions weighted by an unknown mixing proportion, sometimes with the mixing proportion varying as a function of covariates. These same covariates may also be incorporated in the component specific regressions. Nevertheless, support for the mixing proportion is derived from the dependent variable, not the covariate. Mixtures of Poisson regressions, and mixtures of logistic regressions, have been explored (McLachlan and Peel 2000; Wang 1994; Wang et al. 1996). Of these models, only mixtures of Poisson regressions are currently applied in the social sciences (see (Land 2001)).

Mixtures of Poisson regressions depend upon the repeated nature of the dependent variable (counts) for identification (Wang 1994). However, mixtures of logistic regressions require binomial experiments often with may repetitions for identification, particularly if there are covariates (McLachlan and Peel 2000; Wang 1994). Since experimental data are seldom available in the social sciences, it is unlikely that finite mixtures of logistic regressions are widely applicable.

This paper explores the statistical properties of a mixture of logistic regressions, where the mixing of the logistic regressions is given by a finite Gaussian mixture model of a continuous variable, which in the case of logistic regression must be an independent variable. This differs from the general definition of finite mixtures of logistic regressions in that the mixing parameter is no longer an unknown characterized by the dependent variable of the logistic regressions. The mixing of the logistic regressions is instead defined by a finite Gaussian mixture model, which is the marginal distribution (density) of a continuous variable or vector of variables. These same variables may also appear in the regression terms of the model as conventional covariates. We call this a Covariate Density Defined (CDD) finite mixture of logistic regressions. The advantage of the CDD methodology for mixtures of logistic regressions is that it can be applied to Bernoulli observations, that is binomial experiments are not required for the model to be identified. The CDD finite mixtures approach is probably applicable to all densities encountered with GLMs, as well as, life history models, wherever a continuous variable's density can be profitably described by a multi-component finite mixture model. CDD finite mixtures of GLMs could be used, for example, to determine the characteristic differences between the two "subspecies" of crab implied by Pearson's (1894) original application of Gaussian mixture models.

Here the method is applied to the study of the relationship between human birth weight and infant mortality. The fact that birth weight and gestational age distributions are consistently skewed is often interpreted as

evidence that birth cohorts are composed of several heterogeneous subpopulations (Brimblecombe, Ashford and Fryer 1968; Fryer, Hunt and Simons 1984; Karn and Penrose 1951; Wilcox and Russell 1983b). This view is supported statistically by applications of two component finite Gaussian mixture models (Fryer et al. 1984; Gage and Therriault 1998) and mixtures of a Gaussian with one or two non-parametric distributions (Umbach and Wilcox 1996; Wilcox and Russell 1983b) to describe human birth weight distributions. Finite Gaussian mixture models have also been applied to gestational age (Gage 2000) with similar results to those obtained with birth weight. Finally, Gage and others (Gage 2002a; Gage et al. submitted) have applied population based CDD finite mixtures of logistic regressions to determine if the components identified by the finite Gaussian mixture models might differ (i.e. are heterogeneous) with respect to mortality characteristics.

The aim of this paper is to document the statistical properties of CDD - finite mixtures of logistic regressions as a method of modeling the relationship between birth weight and infant mortality. The specific aims are: a) to apply the model to 12 populations by sex and ethnicity born in New York State 1985-88 to establish the range of variation for human populations living in developed environments, b) to identify potential sources of estimation bias in the procedure, c) demonstrate that the parameter estimates are consistent, and d) examine Type I and Type II (1.0 - power) errors of the statistic. Collectively these aims provide a context for interpreting the important characteristics of the relationship between birth weight and infant mortality.

The Model

Gage et al. (submitted) define CDD - finite mixtures of logistic regressions for the Gaussian two-subpopulation case as the joint density of birth weight and occurrence of death:

$$f((x,y);\beta,\theta) = f(y \mid x;\beta,\theta)f(x;\theta) \tag{1}$$

The birth weight density, $f(x;\theta)$, is given by;

$$f(x;\theta=(\theta_1,\theta_2,\pi))=\pi N(x,\theta_1=(\mu_1,\sigma_1^2))+(1-\pi)N(x;\theta_2=(\mu_2,\sigma_2^2)) \qquad (2)$$

with π defined as the proportion of births belonging to the Gaussian density labeled 1, (the mixing proportion), and for $i=1$ to 2, $N(x;\theta_i=(\mu_i,\sigma_i^2))$ being Gaussian densities with mean μ_i and variance σ_i^2 truncated at 0.0

The probability of death conditioned on birth weight is given by:

$$f(y=1\,|\,x;\beta=(\beta^{(1)},\beta^{(2)}),\theta)=q(x;\theta)P(x;\beta^{(1)})+(1-q(x;\theta))P(x;\beta^{(2)}) \qquad (3)$$

where an infant of birth weight x in the ith subpopulation has probability of dying given in quadratic logistic form:

$$P(x;\beta^i)=e^{a_i+b_ix+c_ix^2} \Big/ (1+e^{a_i+b_ix+c_ix^2}) \qquad (4)$$

and $q(x;\theta)$ is the conditional probability that an infant of birth weight x belongs to subpopulation 1. The model's birth weight density form determines that:

$$q(x;\theta)=\pi N(x;\theta_1) \Big/ (\pi N(x;\theta_1)+(1-\pi)N(x;\theta_2)). \qquad (5)$$

The mixing proportion has been transformed to

$$\rho=\text{logit}(\pi)$$

which transforms the 0.0 and 1.0 bounds on π to minus and plus infinity respectively. Subpopulation-and-birth-weight-specific infant mortality is assumed to be U-shaped, hence the quadratic assumption in equation 4. This quadratic form is the parsimonious parameterization of infant mortality in the homogeneous case (Fryer et al. 1984). Altogether there are 11 parameters, 5 defining the mixture, and 6 defining subpopulation-specific mortalities (Table 1).

Table 1 about here.

The CDD-finite mixture of logistic regressions is identified when the finite mixture model, and the individual logistic regressions are identified. In particular it is noted that equation 1 shows the joint density as the product of the marginal birth weight density and a second factor, which is a mixture of logistic

regressions. The mixture birth weight density is identifiable up to assigning which Gaussian density in the mixture model is called subpopulation 1 (McLachlan and Peel 2000). We have followed the convention of assigning subpopulation 1 to the subpopulation accounting for the majority of births (Gage 2002b). For notational convenience this subpopulation is subscripted p, and called the primary subpopulation, and the less numerous subpopulation is subscripted s, and referred to as the secondary subpopulation. Further the logistic terms are identified if the data matrix is full rank. The CDD-finite mixtures of logistic regressions differs from the usual definition of a finite mixture model of logistic regressions, which are not identifiable when the observations are Bernoulli (Wang 1994) (McLachlan and Peel 2000). The difference is that the traditional model defines the mixture as an unknown parameter, where as in the CDD-finite mixtures of logistic regression the mixture is not a parameter, but is a function of the components of the identifiable continuous birth weight mixture model. Thus the CDD-finite mixtures of logistic regression is likely to be applicable in many situations where the traditional model is not.

There are several important but unresolved problems with respect to finite mixture models, choice of parametric distribution (e.g. Gaussian versus log Gaussian etc), and statistically testing the number of components in a mixture. It is likely that these two issues interact, that is the statistically optimum number of components might depend on the assumed parametric densities of the underlying components. Little work has considered alternative parametric component densities in the context of mixture models, however, see Gage (Gage 2002b) for a study of this issue with respect to birth outcomes. On the other hand the development of a statistical test for the number of components in a mixture has received considerable attention in the statistical literature. Since hypotheses concerning the number of components occur on boundaries, the standard likelihood ratio criterion is not appropriate. The best procedure for testing significance is the bootstrap method suggested by McLachlan (1987). Given the

increasing popularity of finite mixture models as a method of cluster analysis, developing a simple statistical test for the number of components is an active area of research (see (McLachlan and Peel 2000) for a review). However, with respect to birth outcomes, previous research using bootstrap methods (Gage 2003), less rigorous penalized likelihood procedures (Gage and Therriault 1998), and theoretical arguments (Gage 2003) all suggest that birth cohorts are composed of at least two Gaussian components.

Further, from the point of view of covariate density defined mixtures of logistic regression the important new issue is whether the covariate defined densities provide useful information concerning heterogeneity in the dependent variable. Hence the hypothesis tests examined here question whether the dynamics of the dependent variable differ between the components, i.e., that is the null hypothesis $a_p=a_s$, $b_p=b_s$ and $c_p=c_s$ given that there are two Gaussian components. In fact it is not necessary that the finite mixture model represent true underlying subpopulation structure. Many arbitrary groupings of individuals have proven useful in practical applications. Within the birth outcomes literature the concept of "low birth weight" is a classic example. It is preferable, however, that the covariate defined densities be theoretically interpretable and approximate the true subpopulation structure.

Data and Methods

Data Sets

The empirical birth outcomes data employed here consists of births to six ethnic groups by sex born in New York State over the period 1985 to 1988 (Table 2). Births from inter-ethnic unions, with missing birth weights, and multiple births are excluded. The data for all ethnic groups are analyzed for the four-year period. Due to their comparatively large sample sizes, the European American birth cohorts are also analyzed disaggregated by year. Sample sizes vary from about 6000 in the case of Asian Americans to over 270,000 in the case of

European Americans over the four-year period. Missing birth weights do not exceed 3/1000 in any of the populations, however, missing birth weights are least common in Asian and African Americans and most common in European American births.

Table 2 about here

Data for simulation studies are generated using the parameter estimates for two observed birth cohorts, that is African American females and European American males, and from 1000 simulated sets of parameters. The 1000 simulated parameter sets are obtained by randomly choosing parameters with a uniform distribution across the range of the observed parameter sets, that is, the observed cohorts listed in Table 2. However, the Asian and African American Hispanic parameter estimates are excluded from the range since these populations are represented by small sample sizes and may incorporate excessive errors. The covariance structure among the 11 parameters is not considered when generating the 1000 parameter sets. Consequently, these parameter sets can and do represent model dynamics well outside the range of variation observed in the birth cohorts listed in Table 2. Each test based on an observed parameter set represents the statistical behavior of the model at a single point in the 11 dimensional parameter space (based on 1000 replicates), while tests based on the 1000 randomly selected parameter sets incorporated 1 replicate at each point and refer to an average statistical behavior across a broad range of conditions.

In all cases, simulated birth cohorts are generated randomly from equation 1 based on the parameter values. Each simulation trial consists of generating a cohort of births randomly from a set of mixture model parameters. The probability of dying is then generated for each simulated birth based on birth weight, subpopulation membership, and the mortality parameters. Deaths to individuals are assigned by generating a random number in the range 0.0 to

1.0 from a uniform distribution. These simulated birth cohorts are then analyzed using the same fitting procedures used to analyze the observed birth cohorts.

Parameter Estimation

The model is fitted using the methods of maximum likelihood with minimization algorithms from the Splus (ms) (Bates and Chambers 1992) and R (nls) (Ihaka and Gentleman 1996) statistical libraries. For each birth cohort several different models are fitted in succession. First a five parameter, two component Gaussian mixture model is fitted to the birth weight distribution. Second a standard homogenous logistic regression is fitted to infant mortality parameterized as a second-degree polynomial of birth weight (Fryer et al. 1984). Third, the population-based mixture of logistic probabilities is fitted to infant mortality where mortality in each component (subpopulation) is parameterized as a separate second-degree polynomial of birth weight. In this case, the five Gaussian mixture model parameters are fixed at the values obtained in step 1, and only the six logistic parameters are allowed to float. Finally, we fit the full model allowing all five mixture parameters and six logistic mortality parameters to float. The parameter estimates from step 1 and step 3 are used as starting values for the final fitting procedure. We have found that the step 1 and step 3 estimates (a two-stage procedure) are usually very similar to estimates obtained from the full model. However, only the full procedure provides true maximum likelihood estimates.

Confidence Intervals

Bias corrected confidence intervals for the parameter estimates, the birth weight specific mixture densities, and the birth weight specific mortality rates are estimated with bootstraps (Staude and Sheather 1990). For each estimate, two bootstraps of 100 iterations each are carried out by sampling with replacement from the observed data set a sample the same size as the observed data set. The bootstrap 95 percentile confidence limits are obtained from the first bootstrap set. Bias is estimated as the difference between the mean of the second bootstrap set

and the estimate obtained from the observed data set. The confidence limits are corrected for bias by adding the bootstrap estimate of bias to the upper and lower confidence limits. Classical confidence intervals for the parameter estimates are also computed from the Hessian using R (Ihaka and Gentleman 1996) for comparison with the bootstrap results.

Statistical Properties of the Parameter Estimates.

Simulation studies are used to determine if the parameter estimates are asymptotically unbiased and consistent. Three test cases are employed; the parameter estimates for the African American females, the European American males, and the sample of 1000 simulated parameter sets. In the case of the observed parameter sets 1000 birth cohorts at sample sizes of 25,000, 50,000 and 100,000 are randomly generated from the known parameters estimates, and then analyzed using the procedures presented above in an attempt to recover the parameters. In the case of the 1000 randomly generated parameter sets the procedures are the same except that only one birth cohort is generated for each of the 1000 simulated parameter sets. Note that the parameters generating the data differ for each trial, where as this is not the case for trials on the observed birth cohorts.

Bias is defined as the difference between the mean of the results obtained from the simulated cohorts and the parameters generating the cohort. Significant bias is identified by a series of t-test. If the mean of the 1000 sets of parameter estimates can not be statistically distinguished from the generating parameters, the method is considered to provide unbiased estimates.

An estimator is considered consistent if the mean square error of the estimates declines to 0.0 as sample size increases to infinity. We present the decline in mean square error of the parameters with respect to sample size, for samples of 25,000, 50,000, and 100,000 using the three standard test cases. In small data sets both observed (see African American Hispanic females below) and simulated, a subpopulation and birth weight-specific mortality curve can be

inverted, that is, rather than a U-shaped mortality curve with a minimum mortality within the range of the data there is an n-shaped mortality curve for one of the subpopulations with a maximum mortality. We refer to this phenomenon as a "flip" in the mortality curve. From a biological point of view U-shaped mortality is expected for a quantitative trait such as birth weight, that is minimum mortality is expected to occur close to the mean of the quantitative trait and higher mortality at very low and very high birth weights. Consequently, n-shaped mortality (a "flip") is not biologically reasonable. It implies lower mortality at the extremes of the birth weight distribution. In any event, this inversion has a large impact on the parameter estimates and hence on the mean square error of the parameter estimates, although the effects on the predicted birth weight specific mortality are generally very small within the range of observed data. Consequently the decline in the frequency of "flips" and the decline in the mean squared error excluding "flips" are presented separately. In addition we have determined if "flips" are more likely in some regions of the 11 dimensional parameter space than others, using the sample of 1000 simulated parameter sets. In this case 50 replicates at each of the 1000 simulated parameter sets are generated for the 25,000 cohort size. A relatively small sample size is used because flips are common only at small sample sizes. The probability of flipping at each of the 1000 points in the 11 dimensional parameter space is then modeled using standard logistic regression. The dependent variable is the flipping phenomena. The independent variables are the values of the generating parameters.

Hypothesis Testing

  Simulation studies are used to examine the use of the standard likelihood ratio criterion for hypothesis testing. Two aspects of hypothesis testing are considered; a) type I error, defined as the probability of rejecting the null hypothesis of homogeneity across components when it is true, and b) power, or 1.0 - type II error, defined as the probability of rejecting the null hypothesis of

homogeneous mortality when mortality differs among the subpopulations. These hypotheses do not include the issue of the number of components in the mixture, since it compares a two-component mixture model with a single logistic mortality model ($a_p=a_s$, $b_p=b_s$, $c_p=c_s$), to a two-component mixture model with separate logistic mortality on each component (equation 1). Since this does not include a test concerning the number of components (a boundry condition), the likelihood ratio criterion is asymptotically chi square (df=3) and should provide accurate type I errors. However, since practical applications are conducted with finite sample sizes, potential questions arise concerning the behavior of the likelihood ratio criterion when π is close to 1.0 the boundry, as well as, the power of an application.

An analysis of type I error similar to that for power described below indicates that the likelihood ratio criterion is generally chi-square. Explorations with various sample sizes and values of π including values as close to 1.0 as 0.995 remain chi-square. There is no suggestion that the likelihood ratio criterion increasingly deviates from chi-square (df=3) at small sample sizes and when π approaches 1.0. These results are available from the corresponding author (tbg), but are not reported further here since we could not find any conditions under which the likelihood ratio criterion failed.

Power is modeled using 10 of the observed, birth cohorts in Table 2 and the model validated against the results obtained with African American females and European American males. African American Hispanic females are omitted due to a "flip" in a mortality curve, (see below). African American female and European American males were excluded since they are used to validate the model. The 1000 simulated parameter sets is not used for this purpose because preliminary analysis indicated that power is close to 100% in these parameter sets except in a very small proportion of the 11 dimensional parameter space where the observed population parameter estimates are located. In the 10 observed cases 50 simulated cohorts of 12,500, 25,000, 50,000, and 100,000 are

generated and power (the dependent variable) was estimated by dividing the number of times the model correctly rejected homogeneous mortality by 50. The independent variables were chosen based on the theoretical likelihood that they might influence power and the about which an investigator might *a priori* be expected to have an opinion with respect to a particular data set. These are: the area difference between the primary and secondary birth weight specific mortality curves[1]; crude death rates for both the primary and secondary distributions; the mixing proportion of the distributions; the difference in the primary and secondary distributions mean birth weights; an indicator of whether the two birth weight specific mortality curves crossed; and sample size. The area difference, the crossing of mortality curves, and the separation between the birth weight distributions are all different aspects of the degree of heterogeneity between subpopulations. Power should increase as heterogeneity increases. The primary and secondary crude death rates, and the mixing proportion, as well as, sample size all influence the magnitude of the number of deaths to be modeled. Power should increase as the number of deaths observed increases.

Power was also empirically estimated for the African American female and European American male test cases. In these cases 100 data sets are generated at 12,000, 25,000, 50,000 and 100,000 sample sizes. Power is estimated as the number of times the model correctly rejects homogeneous mortality (the step 2 fit) divided by the number of trials, in this case 100. These results are described and used to validate the model described above.

Results

The Observed Birth Cohorts

The parameter estimates and confidence intervals obtained with the full model for all 18 populations examined are presented in Tables 3, (the mixture model parameters), 4 and 5, (the primary and secondary mortality parameters respectively). With respect to the five Gaussian mixture model parameters, the

primary component accounts for 87% to 94% of births with the secondary component accounting for the remainder . The primary component consistently has a higher mean birth weight, 3200 to 3546 grams, and smaller standard deviation, 370 to 478 grams, compared to the secondary component. The mean of the secondary component ranges from 2542 to 3179, while the standard deviation varies from 866 to 1158 grams. As a consequence of the larger variance, the secondary distribution accounts for most births in both the lower and upper tails of the birth weight distribution (Figure 1).

<div align="center">

Table 3 about here

Table 4 about here

Table 5 about here

Figure 1 about here

</div>

The subpopulation and birth weight specific infant mortality rates indicate that infant mortality is characteristically U-shaped for both components (Table 5, and Figure 2).   There are several cases where mortality does not increase at the highest birth weights and mortality may be L-shaped rather than U-shaped, i.e., Asian American males and females both primary and secondary mortality, African American male secondary mortality and African American Hispanic male primary mortality. In these cases the confidence intervals for the squared term of the primary mortality polynomial includes 0.0. Nested analyses confirm that these parameters are not significant (Table 6). There is one anomalous case, African American Hispanic females, in which primary mortality is estimated to be n-shaped rather then U-shaped, as indicated by a positive linear (b_1) and negative squared (c_1) term in the mortality polynomial (Figure 3). The bootstrap confidence limits indicate that neither the intercept, linear or squared terms for the primary subpopulation are significantly different from 0.0 for this estimate (Table 4). Nested analysis confirms that an L-shaped mortality model ($c_1 = 0.0$) is parsimonious for African American Hispanic females (Table 6). Figure 4 shows mortality for African American Hispanic females based on this model ($c1=0.0$),

which closely resembles Figure 3 at least over the range where the primary component dominates. The restricted range of birth weights of the primary subpopulation explains why in the full model primary mortality can be n-shaped and still fit the observed mortality data well (Figures 3 and 4). Thus none of the observed data supports an n-shaped birth weight-specific primary or secondary mortality curve. In general, the birth weight specific infant mortality curves are U-shaped (or perhaps L-shaped) for all populations.

<div align="center">

Figure 2 about here

Figure 3 about here

Figure 4 about here

Table 6 about here

</div>

The bootstrapped standard errors of the parameter estimates (Tables 4 and 5) indicate that n-shaped mortality is observed in some simulated samples of the Asian American populations of both sexes, the African American Hispanic male cohort and even the 1985 European American female cohort, in addition to the African Hispanic female cohort. Of these cases nested analysis indicates that mortality is not heterogeneous in three, the Asian (male and female) and the 1985 European (female) cohort (Table 6). Some additional characteristics indicative of the level of heterogeneity in the observed cohorts are presented in Table 7.

<div align="center">

Table 7 about here

</div>

Surprisingly, secondary infant mortality is generally lower at every birth weight compared to primary infant mortality (Figure 2)! Nevertheless, overall (crude) secondary infant mortality is higher then overall crude primary mortality usually by an order of magnitude (Table 7). Thus this system represents an excellent example of Simpson's paradox. In two cases, the mortality curves do cross, that is, for Asian American females and African American Hispanic females (Figure 3 and 4, Table 7). As noted above, the Asian American female cohort can not reject the null hypothesis of homogeneity of mortality. The cross remains for African American Hispanic females even after eliminating

insignificant terms (Figure 4). Both of these results are based on small population sizes and may be unreliable.

As a result of the characteristic patterns of subpopulation and birth weight specific mortality and the dynamics of the mixture, total mortality, that is combined across both subpopulations, is not a simple U-shape (Figures 2, 3 and 4). In particular, there is a shoulder at about 2500 grams and often a second shoulder at 5500 grams (Figures 2,3, 4). The shoulders are due to the dynamics of the mixture and the differences between the mortality curves. The shoulders occur at birth weights where transitions from predominately secondary births to predominately primary births and then back again to predominately secondary births take place.

Bootstrapped confidence intervals on the mortality curves vary across birth weights (Figure 5). The intervals tend to be larger where the shoulders occur, that is, where the transitions from predominately secondary to predominately primary occur. The decrease in confidence in these regions is probably a result of the decreased classificatory power of the mixture model at birth weights where the mixture is approximately 50% primary and 50% secondary.   The relatively small number of births at heavier birth weights also contributes to the large confidence intervals at large birth weights.   As a result it is not clear if these upper shoulders are real or artifacts of the model specification (quadratic functions of birth weight).

Figure 5 about here.

The bootstrapped confidence intervals for the parameter estimates (Tables 2 and 3) suggest that the Hessian can considerably underestimate the true confidence intervals.   The ratio of Hessian to bootstrapped confidence interval lengths for two cases, African American females and European American males, are presented in Table 8. The mean ratio across all 11 parameters is 0.96 in the case of African American females but only 0.80 in the case of European American males. The birth weight mixture model parameter standard errors estimated

from the Hessian and the bootstrap are reasonably similar. However, the mortality logistic regression parameter standard errors are considerably smaller when estimated from the Hessian. These results suggest that the likelihood surface may not be well approximated as a quadratic at the solution. The bootstrap confidence intervals are preferred, since they make no assumption about the curvature of the likelihood at its maximal value.

Table 8 about here.

The Statistical Properties of the Parameter Estimates

The fitting procedure provides unbiased estimates of the parameter values at least with samples of 25,000 to 100,000 (Table 9). In general, the estimated bias is small and/or declines with sample size. None of the estimated biases approach statistical significance (the largest value of t is 0.63).

Table 9 about here

The mean square error of the parameter estimates exclusive of "flips" declines with sample size, indicating that the parameter estimates are consistent (Table 10). The mean square error of the estimates declines fastest for the African American female data and slowest for the 1000 simulated data sets. Two of the comparisons for the simulated parameter sets do not decline; $\pi.$ (25,000/50,000) and σ_1 (50,000/100,000). This is not surprising given that there are 66 comparisons altogether. Moreover, both of these cases decline across the 25,000/100,000 comparison. In general mean square error declines with sample size in all cases indicating the procedure is asymptotically consistent.

Table 10 about here

The results presented in Table 10 do not include cases in which the fits displayed n-shaped mortality.   These were simply omitted from the calculations. In all cases, the parameters generating these simulated data represent heterogeneous U-shaped mortality curves. Nevertheless, n-shaped mortality curves were observed in the European American male case and in the 1000 simulated parameter sets (Table 11). No cases of flips are observed with the

African American female simulations at any sample size. Clearly this problem diminishes as sample size increases. Further the differences between the African American female and European American male results suggest that the frequency of "flips" varies across the 11 dimensional parameter space. Inclusion of the n-shaped mortality results in Table 10 would simply make the decline in mean square error with sample size more dramatic. Again the procedure appears to be asymptotically consistent.

<div align="center">Table 11 about here</div>

To understand the variation in "flipping" across the 11 dimensional parameter space, 50 simulation trials are carried out for each of the 1000 simulated data sets at the 25,000 cohort size. This sample size was used because these issues are most extreme at smaller sample sizes. The results were split into 2 groups: n-shaped mortality in the primary subpopulation and flips in the secondary subpopulation. The probability of flipping was modeled with logistic regression using the simulated data's 11 generating parameters as the independent variables. The logistic regression model for predicting flips in the primary subpopulation and birth weight specific mortality is presented in Table 12. The results suggest that all 11 parameters significantly influence the probability of an n-shaped mortality curve, but the mortality parameters are the most influential. In particular, as secondary mortality increases (positive coefficients) and primary mortality decreases (negative coefficients) the probability of a primary flip increases. This suggests that as the two birth weight specific mortality curves come closer together (that is, lower heterogeneity) the primary subpopulations birth weight specific mortality curve is more likely to flip.

<div align="center">Table 12 about here</div>

The logistic regression model for secondary flips is presented in Table 13. In this case, the variables that have the greatest influence on the probability of a secondary flip are the proportion of births attributed to the secondary

distribution, the mean of the secondary distribution and all 3 secondary mortality coefficients. This suggests that secondary flips are affected mostly by characteristics of the secondary distribution. As the number of births in the secondary distribution decreases and/or secondary mortality decreases the probability of a secondary flip increases. This suggests that the secondary flipping phenomenon is solely a sample size problem, i.e. the number of infant deaths.

<div align="center">Table 13 about here</div>

Power

Samples of 50,000 births appear to be necessary to *insure* powers of 80% to reject the homogeneous null hypothesis when it is false (Table 14). For example, African American females require sample sizes of only 12,500, while European American males need sample sizes of at least 50,000 births to achieve 80% power. In both cases power approaches 100% for samples of 100,000. Clearly, relatively large sample sizes are necessary, however, the characteristics of the population are also an important determinant of power.

<div align="center">Table 14 about here</div>

To explore the characteristics that effect power, and to provide a method of estimating power for future applications we have estimated power for the ten remaining independent data sets described in Table 2 , excluding African American females, European American males and Hispanic African American females. The Hispanic African American female cohort is eliminated due to the n-shaped primary mortality curve, which is theoretically anomalous and probably due to small sample size. The empirical power estimates for these ten populations are consistent with the findings presented earlier concerning heterogeneity in mortality. The null hypothesis of homogeneity could not be rejected in the cases of Asian females and males and European American females in 1985 (Table 6). The estimates presented in Table 15 indicate that power is less then 50% in all three cases given the sample sizes in Table 2. The covariates for

the analysis of power, except for sample size, are all logically related to the apparent heterogeneity of the birth cohort and are presented in Table 7.

Table 15 about here

Logistic regression analysis of the data in Table 15 with the data in Table 7 as covariates, suggest that in addition to sample size, power is a function of the heterogeneity between the subpopulations (Table 16). Power increases as five of the covariates increase. The most important factor is of course sample size. A higher overall primary crude death rate also increases power. But, the other measures of the magnitude of deaths modeled, secondary crude death rates and the mixing proportion do not appear to influence power. On the other hand all of the indicators of heterogeneity influence power. The area between the mortality curves[1] is almost as important as sample size. If the birth weight and subpopulation specific mortality curves cross, indicating different birth weight specific dynamics, power tends to increase. Similarly the difference in birth weight means, that is, the greater the separation between the subpopulation birth weight densities, increases power. Thus increased heterogeneity in mortality, and birth weight density between the subpopulations is associated with greater power.

Table 16 about here

Validation of this model using the estimates of power obtained with simulation techniques for African American females and European American males (Table 13), two cohorts excluded from the logistic regression sample, is presented in Table 17. The African American female cohort has a greater area difference between the subpopulation specific mortality curves (by a factor of 1.2), a greater difference in birth weight means (by a factor of 1.2) and a greater crude primary death rate (by a factor of 1.8) than did the European American male cohort (Table 7). The mortality curves did not cross in either data set (Figure 2). As a result the predicted power for African American females is higher than for European American males. This is consistent with the estimates

of power for these same populations based on simulation trials. In general the model does an excellent job predicting power for the African female and European male data sets. Power is typically slightly underestimated making the result appropriately conservative. The notable exception is European American males at 100,000, which is over estimated. In this case, however, the power estimate is above 90% using either prediction.

Table 17 about here

Discussion

Perhaps one limitation of CDD – finite mixtures of logistic regressions in the context of analyzing infant mortality in low mortality populations is that relatively large sample sizes are required to reject the null hypothesis of homogeneity when it is incorrect. At least in one case given above, samples greater than 50,000 are necessary to insure reasonable power. This is partly because infant mortality is a relatively rare event, at least in the developed areas of the world such as New York State, and because the degree of heterogeneity between the subpopulations is not necessarily large both in terms of mortality and separation of birth weight densities. On the other hand, sample sizes as small as 12,000 are sufficient in other cases with greater heterogeneity, such as African American females. In any event in low mortality populations large data sets are generally available concerning birth outcomes. Considerably smaller samples may be sufficient in high mortality populations where large data sets are harder to obtain. A model for predicting power is provided to guide future investigations.

In practical applications of CDD-finite mixtures of logistic regressions to birth outcomes data, the theoretical principles of maximum likelihood appear to apply. Type I error does not appear to be problematic even when π approaches 1.0, although the likelihood ratio criterion fails theoretically at the boundry, $\pi=1.0$. Finely, the parameter estimates are asymptotically consistent and unbiased.

The nested analysis and simulation studies both indicate that n-shaped mortality curves ("flips") are a function of small samples and the limited range of primary subpopulation birth weights. First, the analysis of the observed population that "flipped" is associated with non-significant parameter estimates. Flips do not occur with parsimonious models where non-significant parameters have been eliminated. Second, simulation studies, where the mortality pattern is known to be U-shaped, also occasionally return an n-shaped pattern. In all cases that "flip", observed and simulated, the declining right side of an n-shaped second degree polynomial appears to fit marginally (but not significantly) better then a U-shaped model. Finally, the "flips" are unlikely to confuse a knowledgeable investigator at least in applications to quantitative biological traits, since n-shaped mortality curves make no biological sense. Thus there is good reason to believe that the n-shaped mortality pattern estimated for African American Hispanic females is a statistical artifact. It is possible that the flipping behavior might be eliminated by an alternative specification of the subpopulation and birth weight specific mortality curves, substituting something less flexible than a second degree polynomial.

Considering available sample sizes for the observed populations and power, the CDD finite mixture of logistic regressions model suggests that birth cohorts typically have the following characteristics: a) the birth cohort is composed of at least two subpopulations heterogeneous with respect to infant mortality, b) the primary subpopulation (accounting for the majority of infants) has a higher mean and lower standard deviation then the secondary subpopulation, as a result of the high standard deviation the secondary subpopulation accounts for the majority of births at both low and high birth weights, c) the secondary subpopulation is the population at highest overall risk, but the birth weight specific mortality of the secondary subpopulation is generally lower at most birth weights, d) both primary and secondary birth weight specific mortality is U or perhaps L shaped, and d) the birth weight

specific total infant mortality curve has a shoulder around 2000 grams. The observed populations that are exceptions to these characteristics all have very low statistical power. These results are consistent with earlier results concerning birth weight distributions (Fryer et al. 1984; Gage and Therriault 1998) and infant mortality (Gage 2002a).

It is hypothesized that the primary subpopulation represents a "normal" fetal development group, while the secondary subpopulation represents fetuses that have been disturbed or compromised during fetal development (Fryer et al. 1984; Gage and Therriault 1998). The lower birth weight specific infant mortality of the secondary subpopulation may be due to higher fetal losses to this group who are consequently more robust than births at the same birth weight in the primary subpopulation (Gage 2002a). Further analysis and comparison of the African and European birth cohorts presented above indicate that the pediatric paradox, the observation that low birth weight African American births have lower mortality than low birth weight European American and presumably socially advantaged births(Gage et al. submitted), is entirely due to the secondary subpopulation. African American primary births have higher birth weight specific mortality compared to European American primary births. On the other hand, African American secondary births have lower birth weight specific mortality than European American secondary births (Gage et al. submitted) possibly as a result of higher fetal loss rates among African Americans. Thus the lower mortality at low birth weights among African Americans might be due to higher levels of stress resulting in higher fetal losses in the disadvantaged population, resolving but not identifying the ultimate cause of the paradox (Gage et al. submitted).

The CDD-finite mixtures of logistic regressions can be elaborated in a number of ways. First, other indicators of heterogeneity could be studied, for example, in the context of birth outcomes, gestational age (Gage 2000). Second, multivariate mixtures can be substituted for univariate mixture models. In the

case of birth outcomes, a birth weight by gestational age mixture would be a reasonable extension (Gage 2003). Third, finite mixture models need not be restricted to Gaussian mixture models. Other parametric specifications could also be examined (Gage 2002b). Fourth, the application is not necessarily limited to logistic regression, but could be generalized to other types of population based regression analysis as well. Finally, covariates, can be introduced into the mixture and logistic terms of the model, e.g. in the context of birth outcomes, maternal age, parity or SES etc. could be introduced into the mixture or logistic models or both. Covariates on the characteristics of the birth weight mixture model influence infant mortality indirectly through birth weight (Gage 2003). Additional covariates (other than birth weight) incorporated into the logistic probabilities influence infant mortality independently of birth weight (i.e. directly) and/or interact with birth weight. In theory by placing the same covariate in the mixture and logistic terms the direct and indirect effects of a covariate on infant mortality can be disentangled, fully operationalizing the proximate determinants model of infant mortality (Eberstein 1989). Further, the covariates could have differential and even opposite effects on the two subpopulations. Thus conventional analyses assuming homogeneity of birth cohorts may underestimate or even completely overlook the effects of important covariates. An analysis with covariates might elucidate the causes of the differences between African and European American birth weight distributions and infant mortality rates. Conventional analyses have been unable to resolve these differences (Costa 2003; Wilcox and Russell 1990).

CDD – finite mixtures of general linear models provide a new method of controlling for hidden heterogeneity. It differs from other models of "hidden" heterogeneity (McLachlan and Peel 2000; Wang 1994) (GLMs) and (Heckman and Singer 1982; Vaupel, Manton and Stallard 1979) (failure time models), which the mixing parameter is an unknown. CDD – finite mixtures of general linear models, on the other hand, attempt to obtain information about "hidden"

heterogeneity from a finite mixture model. Unlike the previous methods, CDD base models cannot correct for sources of heterogeneity that are not reflected in the chosen covariates marginal distribution. The disadvantage of this approach is that some unmeasured heterogeneity may remain in the analysis. The advantage is that the causes and consequences of the heterogeneity identified can be explored. For example, in the birth weight example presented above, the finding that there are two subpopulations in the birth cohort immediately raises several questions, do these subpopulations differ with respect to mortality? What are the characteristics of the two subpopulations? etc. For example, the CDD-finite mixtures of GLMS could be used to examine the hypothesis that the secondary subpopulation is a product of disturbed fetal development by determining if adverse conditions of pregnancy are associated with the secondary subpopulation, etc. On the other hand, with the generic models the sources of "hidden" heterogeneity are controlled for but remain hidden. Thus the CDD finite mixtures of general linear models is less general, but potentially more informative because it naturally leads the investigator to ask questions concerning the source of the heterogeneity, as well as, providing a methodology for identifying the sources. It will be useful even with finite mixtures of Poisson regressions where the generic approach is also possible. As a result, currently "hidden" heterogeneity may eventually become directly observable and understood.

The CDD approach has applicability wherever finite mixture models are useful. The utility of the method for studying infant mortality with the addition of covariates that are "proximate determinants model of infant mortality" (Eberstein 1989; Menken 1987), is clear based on the results presented above. The same sort of analysis is relevant to examination of the fetal origins hypothesis, i.e., that conditions, (disturbances) during fetal development may cause adult sequelae, for example, the association of low birth weight with heart disease in adults (Godfrey and Barker 2000). Further, given the explosive growth of finite

mixture models as a method of cluster analysis (McLachlan and Basford 1988; McLachlan and Peel 2000), CDD finite mixtures are likely to identify potentially heterogeneous components in a broad range of variables and hence be useful in a broad range of applications. The covariate density defined finite mixture method could potentially double the information available for analyses, since the same covariate can be used to identify heterogeneity "hidden" by its density distribution, as well as, serve as a conventional covariate measuring the level of the variable.

Conclusions

Maximum likelihood estimates of Covariate Density Defined Finite Mixtures of Logistic Regressions applied to birth weight and infant mortality are asymptotically consistent and unbiased. In low mortality settings (such as New York State) relatively large sample sizes (>50,000) are needed to insure sufficient power. Fortunately, in low mortality settings large samples are readily available. Considerably smaller sample sizes may be sufficient in high mortality settings, where large data sets may not be available.

The analyses indicate that the relationship between birth weight and infant mortality has the following characteristics, a) birth cohorts are composed of two or more subpopulations, b) the majority subpopulation has a higher mean and lower variance compared to the minority subpopulation, and so the minority subpopulation accounts for most births at both extremes of the birth weight distribution, c) the minority subpopulation has higher crude death rates but lower birth weight specific mortality at every birth weight compared to the majority subpopulation, and d) total birth weight specific infant mortality is not a simple U-shape but has a shoulder at around 2000 grams and perhaps a second at higher birth weights..

Finally, covariate density defined finite mixtures of logistic regressions is one method of correcting for sources of "hidden" heterogeneity. Given the increasing use of finite mixture modeling for cluster analysis, it is likely that

potentially "hidden" heterogeneity may be identifiable from the marginal distribution of a wide variety of covariates. This could double the information available to these analyses, since the covariates may also be incorporated as conventional covariates. Consequently, population based parametric mixtures of logistic regressions, as well as, parametric mixtures of other types of regression should have broad statistical applicability.

Footnotes

1. Area difference between the primary and secondary birth weight specified mortality curves (shown in Figure 2) is a relative measure. In this study the area differences are measured between 50 and 6000 grams and mortality is a rate per 1,000 births. This standardizes the measurements to have a maximum value of 5,950,000. The mean value of the 13 independent data sets used in the power analyses is 1,072,175 with a standard deviation of 222,380.6. Estimates for each of the observed birth cohorts are presented in Table 5a.

References

Bates, D.M.and J.M. Chambers. 1992. "Nonlinear Models." Pp. 421-454 in *Statistical Models in S*, edited by J.M. Chambers and T.J. Hastie. Pacific Grove, CA: Wadsworth and Brooks.

Brimblecombe, F.S.W., J.R. Ashford, and J.G. Fryer. 1968. "Significance of low birthweight in perinatal mortality: a study of variations within England and Wales." *Br. J. Prev Soc Med* 22:27-35.

Costa, D.L. 2003. "Race and pregnancy outcomes in the twentieth century: A long-term comparison." National Bureau of Economic Research, Working Paper No. W9593.

Eberstein, I.W. 1989. "Demographic researh on infant mortality." *Sociological Forum* 4(3):409-422.

Fryer, J.G., R.G. Hunt, and A.M. Simons. 1984. "Biostatistical considerations: The case for using models." Pp. 9-30 in *Prevention of Perinatal Mortality and Morbidity*, edited by F. Falkner. Basel: Karger.

Gage, T.B. 2000. "Variability of gestational age distributions by sex and ethnicity; An analysis using mixture models." *American Journal of Human Biology* 12:181-191.

—. 2002a. "Birthweight-specific infant and neonatal mortality:The influence of heterogeneity in the birth cohort." *Human Biology* 74:165-184.

—. 2002b. "Modeling birthweight and gestational age distributions:Additive vs. multiplicative processes." *American Journal of Human Biology* 14:1-7.

—. 2003. "Classification of births by birth weight and gestational age: An application of multivariate mixture models." *Annals of Human Biology* 30(5):589-604.

Gage, T.B., M.J. Bauer, N. Heffner, and H. Stratton. submitted. "The pediatric paradox: Heterogeneity in the birth cohort." *Human Biology*.

Gage, T.B.and G. Therriault. 1998. "Variability of Birth-Weight Distributions by Sex and Ethnicity: Analysis Using Mixture Models." *Human Biology* 70(3):517-534.

Godfrey, K.M.and D.J.P. Barker. 2000. "Fetal nutrition and adult disease." *American Journal of Clinical Nutrition* 71(suppl):1344s-1352s.

Heckman, J.and J.J. Singer. 1982. "Population heterogeneity in demographic models." Pp. 567-599 in *Multidimensional Mathematical Demography*, edited by K. Land and A. Rogers. New York: Academic Press.

Ihaka, R.and R. Gentleman. 1996. "R: A language for data analysis and graphics." *Journal of Computational and Graphical Statistics* 5(3):299-314.

Karn, M.N.and L.S. Penrose. 1951. "Birthweight and gestation time in relation to maternal age, parity and infant survival." *Ann. Eugen* 16:147-160.

Land, K.C. 2001. "Introduction to the special issue on finite mixture models." *Sociological Methods and Research* 29(3):275-281.

McLachlan, G.J.and K.E. Basford. 1988. *Mixture Models: Inference and Applications to Clustering*. New York: Marcel Dekker.

McLachlan, G.J.and D. Peel. 2000. *Finite Mixture Models*. New York: John Wiley and Sons INC.

Menken, J. 1987. "Proximate determinants of fertility and mortality: A review of recent findings." *Sociological Forum* 2:697-717.

Staude, R.and S. Sheather. 1990. *Robuts Estimation and Testing*. New York: Wiley.

Umbach, D.M.and A.J. Wilcox. 1996. "A technique for measuring epidemiologically useful features of birthweight distributions." *Statistics in Medicine* 15:1333-1348.

Vaupel, J.W., K.G. Manton, and E. Stallard. 1979. "The impact of heterogeneity in individual frailty on the dynamics of mortality." *Demography* 16:439-454.

Wang, P. 1994. "Mixed Regression Models for Discrete Data." University of British Columbia.

Wang, P., M.L. Puterman, I. Cockburn, and N.D. Le. 1996. "Mixed poisson regression models with covariate dependent rates." *Biometrics* 52:381-400.

Wilcox, A.and I. Russell. 1990. "Why small black infants have a lower mortality rate than small white infants: The case for population-specific standards for birth weight." *J. of Ped.* 116:7-10.

Wilcox, A.J.and I.T. Russell. 1983b. "Birthweight and perinatal mortality: I. On the frequency distribution of birthweight." *International Journal of Epidemiology* 12(3):314-318.

Figure Captions

Figure 1. The mixture model fitted to African American female births (1985-88).
The results are qualitatively similar for the other populations examined. The
solid line represents the total density, while the short dashed line is the density
of subpopulation 1 and the long short dashed line is the density of subpopulation
2. The rug plot displays the density of the original data.

Figure 2. The mortality model fitted to African American female births (1985-88).
The results are similar for the other populations (see text for qualifications). The
solid line represents the total death rate, while the short dashed line is the death
rate of subpopulation 1 and the long short dashed line is the death rate of
subpopulation 2. The rug plot displays the density of births.

Figure 3. The full mortality model fitted to African American Hispanic female
births (1985-88). The solid line represents the total death rate, while the short
dashed line is the death rate of subpopulation 1 and the long short dashed line is
the death rate of subpopulation 2. The n-shaped birth weight specific mortality
cure for subpopulation 1 is biologically unrealistic.

Figure 4. A reduced mortality model ($c1=0.0$) fitted to African American
Hispanic female births (1985-88). The solid line represents the total death rate,
while the short dashed line is the death rate of subpopulation 1 and the long
short dashed line is the death rate of subpopulation 2. Subpopulation 1 mortality
declines linearly with birth weight (the graph shows this on a log scale) and is
biologically more realistic, although in general quantitative traits are expected to
display high mortality at both high and low levels of the trait.

Figure 5. (actually 7 in ppt). The mortality model with 95% confidence intervals. The solid line represents the total death rate (same as Figure 2). The dashed lines represent the 95% confidence limits for the total death rate. The rug plot displays the density of all births.

Table 1 Definitions of the model parameters

Symbol	Definition
	Mixture Parameters
π	Mixing proportion (% primary subpopulation)
μ_1	Mean primary subpopulation
μ_2	Mean secondary subpopulation
σ_1	Standard Deviation primary subpopulation
σ_2	Standard Deviation secondary subpopulation
	Mortality Parameters (a, b, and c are coefficients of a second degree polynomial)
a_1	Constant primary subpopulation
b_1	Linear term primary subpopulation
c_1	Squared term primary subpopulation
a_2	Constant term secondary subpopulation
b_2	Linear term secondary subpopulation
c_2	Squared term secondary subpopulation

Table 2 Summary Statistics for the Observed NYS Birth Cohorts 1985-88

Birth Cohort	Births	NA*	Mean &	StDev.&	Skewness	Deaths	Mortality*
As. Am. f.^	6118	0.65	3220	449	0.51	27	4.41
As Am. m.^	6656	0.15	3316	460	0.37	27	4.06
His. Af. Am. f.^	7395	1.89	3216	579	0.87	68	9.20
His.Af. Am. m.^	7725	2.72	3338	598	0.85	96	12.43
Af. Am. f.	54968	1.27	3119	611	0.88	690	12.55
Af. Am. m.	57449	1.64	3240	627	0.96	848	14.76
His. Am. f.	68708	2.42	3263	543	0.72	479	6.97
His. Am. m.	72411	2.51	3376	578	0.67	586	8.09
Eur Am. f.	255516	1.99	3375	524	0.55	1294	5.06
Eur Am. f. 1985	62863	1.30	3369	519	0.54	332	5.28
Eur. Am. f. 1986	62379	1.73	3376	523	0.54	314	5.03
Eur. Am. f. 1987	63635	2.44	3375	524	0.59	296	4.65
Eur. Am f. 1988	66639	2.45	3379	529	0.52	352	5.28
Eur. Am. m.	270189	2.09	3507	557	0.61	1734	6.42
Eur Am. m. 1985	67005	1.42	3502	554	0.59	464	6.92
Eur Am. m. 1986	65733	1.83	3510	555	0.59	417	6.34
Eur. Am. m. 1987	66835	2.39	3509	557	0.64	407	6.09
Eur. Am. m. 1988	70616	2.70	3509	563	0.61	446	6.32

As.=Asian, Af. =African, His. = Hispanic, Eur.= European, Am. = American, f. = female, and m. = male.

NA = births missing birth weight data

* = rate / 1000 births

& = birth weight

^=excluded from some analyses due to small numbers of births and deaths

Table 3 The Mixture Parameter Estimates and 95% Confidence Limits

Birth Cohort	π	(LCL – UCL)	μ_1	(LCL – UCL)	μ_2	(LCL – UCL)	σ_1	(LCL – UCL)	σ_2	(LCL – UCL)
As. Am. f.	0.90	(0.86 - 0.94)	3236	(3222 – 3248)	3065	(2888 - 3175)	370	(355 – 384)	888	(776 – 1014)
As Am. m.	0.90	(0.81 - 0.94)	3332	(3318 – 3343)	3179	(3057 – 3301)	386	(363 – 401)	866	(729 – 952)
His. Af. Am. f.	0.88	(0.85 - 0.91)	3283	(3271 – 3292)	2693	(2531 – 2850)	442	(429 – 458)	1077	(1003 – 1153)
His.Af. Am. m.	0.91	(0.88 - 0.93)	3400	(3387 – 3412)	2706	(2481 – 2871)	473	(458 – 487)	1163	(1073 – 1222)
Af. Am. f.	0.87	(0.86 - 0.88)	3200	(3196 – 3205)	2542	(2478 – 2594)	455	(449 – 461)	1110	(1086 – 1140)
Af. Am. m.	0.88	(0.87 - 0.89)	3328	(3323 – 3332)	2577	(2511 – 2626)	465	(460 – 470)	1138	(1112 – 1162)
His. Am. f.	0.93	(0.92 - 0.93)	3307	(3302 – 3311)	2690	(2634 – 2752)	451	(447 – 455)	1079	(1044 – 1110)
His. Am. m.	0.90	(0.89 - 0.91)	3429	(3424 – 3433)	2906	(2845 – 2971)	467	(462 – 473)	1074	(1041 – 1108)
Eur Am. f.	0.94	(0.94 - 0.95)	3403	(3401 – 3404)	2916	(2870 – 2948)	451	(450 – 454)	1120	(1096 – 1142)
Eur Am. f. 1985	0.94	(0.94 - 0.95)	3396	(3392 – 3400)	2908	(2827 – 2989)	449	(444 – 454)	1115	(1063 – 1164)
Eur. Am. f. 1986	0.94	(0.93 - 0.95)	3402	(3399 – 3406)	2946	(2877 – 3029)	452	(446 – 455)	1115	(1057 – 1163)
Eur. Am. f. 1987	0.94	(0.94 - 0.95)	3405	(3401 – 3408)	2864	(2784 – 2928)	453	(450 – 457)	1115	(1078 – 1160)
Eur. Am f. 1988	0.94	(0.93 - 0.95)	3407	(3404 – 3412)	2939	(2875 – 3003)	452	(449 – 455)	1131	(1091 – 1185)
Eur. Am. m.	0.94	(0.93 - 0.94)	3543	(3541 – 3545)	2988	(2952 – 3028)	476	(474 – 478)	1129	(1108 – 1151)
Eur Am. m. 1985	0.93	(0.92 - 0.94)	3537	(3532 – 3540)	3041	(2963 – 3106)	472	(467 – 476)	1094	(1056 – 1134)
Eur. Am. m. 1986	0.93	(0.93 - 0.94)	3545	(3541 – 3548)	3005	(2929 – 3073)	476	(471 – 480)	1118	(1071 – 1158)
Eur. Am. m. 1987	0.94	(0.93 - 0.95)	3546	(3542 – 3550)	2907	(2838 – 2983)	478	(474 – 482)	1141	(1109 – 1187)
Eur. Am. m. 1988	0.93	(0.93 - 0.94)	3544	(3540 – 3548)	2992	(2906 – 3051)	478	(473 – 482)	1159	(1114 – 1206)

As.=Asian, Af. =African, His. = Hispanic, Eur.= European, Am. = American, f. = female, and m. = male.

LCL=lower 95% confidence limit, UCL=upper 95% confidence limit

Table 4. The Mortality Parameter Estimates and 95% Confidence Limits, Primary Subpopulation

Birth Cohort	a_1 (LCL UCL)	b_1 (LCL UCL)	c_1 (LCL UCL)
As. Am. f.	11.0 (-535.8 22.8)[1]	-0.011 (-0.018 0.356)[ns]	1.70e-6 (-6.16e-5 2.77e-6)[ns]
As Am. m.	16.1 (-8124.1 25.0)[ns]	-0.012 (-0.018 3.886)[ns]	1.67e-6 (-6.37e-4 2.46e-6)[ns]
His. Af. Am. f.	-2.4 (-42.8 8.7)[ns]	0.001 (-0.007 0.031)[ns]	-5.18e-7 (-5.32e-6 7.00e-7)[ns]
His.Af. Am. m.	16.4 (-7.8 20.8)[ns]	-0.012 (-0.015 0.001)[ns]	1.72e-6 (-2.45e-7 2.13e-6)[ns]
Af. Am. f.	9.9 (6.5 13.3)	-0.009 (-0.011 -0.007)	1.19e-6 (8.66e-7 1.53e-6)
Af. Am. m.	9.7 (7.2 11.7)	-0.008 (-0.010 -0.007)	1.12e-6 (8.68e-7 1.31e-6)
His. Am. f.	10.5 (6.7 13.0)	-0.009 (-0.011 -0.007)	1.25e-6 (8.29e-7 1.50e-6)
His. Am. m.	12.7 (7.8 14.8)	-0.010 (-0.011 -0.007)	1.33e-6 (8.67e-7 1.51e-6)
Eur Am. f.	13.7 (11.5 15.3)	-0.011 (-0.012 -0.010)	1.47e-6 (1.27e-6 1.63e-6)
Eur Am. f. 1985	10.5 (-16.0 14.2)[ns]	-0.009 (-0.011 0.007)[ns]	1.19e-6 (-1.27e-6 1.57e-6)[ns]
Eur. Am. f. 1986	14.1 (9.4 16.1)	-0.011 (-0.013 -0.008)	1.53e-6 (1.08e-6 1.75e-6)
Eur. Am. f. 1987	14.1 (7.2 17.5)	-0.011 (-0.013 -0.008)	1.50e-6 (9.77e-7 1.80e-6)
Eur. Am f. 1988	14.7 (10.4 17.1)	-0.011 (-0.013 -0.009)	1.55e-6 (1.06e-6 1.78e-6)
Eur. Am. m.	11.7 (8.9 13.6)	-0.009 (-0.010 -0.008)	1.18e-6 (9.37e-7 1.33e-6)
Eur Am. m. 1985	12.9 (9.4 15.8)	-0.010 (-0.012 -0.008)	1.29e-6 (1.01e-6 1.55e-6)
Eur. Am. m. 1986	11.2 (3.6 14.3)	-0.009 (-0.011 -0.005)	1.09e-6 (4.82e-7 1.33e-6)
Eur. Am. m. 1987	10.6 (5.9 14.3)	-0.009 (-0.011 -0.006)	1.09e-6 (6.78e-7 1.41e-6)
Eur. Am. m. 1988	11.0 (2.5 14.3)	-0.009 (-0.011 -0.004)	1.15e-6 (5.69e-7 1.41e-6)

As.=Asian, Af. =African, His. = Hispanic, Eur.= European, Am. = American, f. = female, and m. = male.

LCL=lower 95% confidence limit, UCL=upper 95% confidence limit

[ns] = not significantly different from zero based on the bootstrapped confidence limits

Table 5. The Mortality Parameter Estimates and 95% Confidence Limits, Secondary Subpopulation

Birth Cohort	a_2 (LCL UCL)	b_2 (LCL UCL)	c_2 (LCL UCL)
As. Am. f.	2.67 (-6.20 7.71)[ns]	-0.004 (-0.011 0.008)[ns]	4.64e-7 (-6.32e-6 1.48e-6)[ns]
As Am. m.	5.30 (-3.59 388.78)[ns]	-0.007 (-0.398 0.025)[ns]	7.82e-7 (-2.07e-5 4.17e-5)[ns]
His. Af. Am. f.	3.96 (0.89 5.65)	-0.006 (-0.008 0.001)[ns]	8.65e-7 (-2.72e-6 1.23e-6) [ns]
His.Af. Am. m.	4.66 (2.67 7.88)	-0.007 (-0.011 - 0.004)	9.19e-7 (4.84e-7 1.56e-6)
Af. Am. f.	3.84 (3.29 4.49)	-0.006 (-0.007 - 0.005)	8.98e-7 (7.29e-7 1.08e-6)
Af. Am. m.	4.42 (3.70 5.07)	-0.006 (-0.007 - 0.005)	7.11e-7 (-1.30e-7 8.30e-7)[ns]
His. Am. f.	3.87 (3.02 4.82)	-0.006 (-0.007 - 0.005)	8.64e-7 (5.41e-7 1.02e-6)
His. Am. m.	3.70 (2.93 4.64)	-0.005 (-0.006 - 0.004)	5.73e-7 (4.56e-7 7.35e-7)
Eur Am. f.	3.39 (2.95 3.86)	-0.005 (-0.006 - 0.005)	5.46e-7 (4.87e-7 6.30e-7)
Eur Am. f. 1985	3.15 (1.80 4.56)	-0.005 (-0.007 - 0.003)	5.15e-7 (3.36e-7 8.13e-7)
Eur. Am. f. 1986	2.89 (1.88 3.61)	-0.005 (-0.005 - 0.003)	4.42e-7 (-5.43e-7 5.90e-7) [ns]
Eur. Am. f. 1987	3.31 (2.41 4.59)	-0.005 (-0.007 - 0.004)	5.34e-7 (4.00e-7 8.32e-7)
Eur. Am f. 1988	4.18 (3.01 5.83)	-0.006 (-0.009 - 0.005)	6.72e-7 (5.29e-7 9.51e-7)
Eur. Am. m.	3.89 (3.48 4.30)	-0.005 (-0.006 - 0.005)	5.73e-7 (4.94e-7 6.78e-7)
Eur Am. m. 1985	4.04 (3.29 5.13)	-0.005 (-0.006 - 0.004)	5.29e-7 (3.42e-7 7.50e-7)
Eur. Am. m. 1986	3.87 (2.79 4.71)	-0.005 (-0.006 - 0.004)	5.24e-7 (3.05e-7 6.56e-7)
Eur. Am. m. 1987	4.11 (3.41 4.97)	-0.006 (-0.007 - 0.005)	6.98e-7 (5.04e-7 8.36e-7)
Eur. Am. m. 1988	3.53 (2.77 4.66)	-0.005 (-0.006 - 0.004)	5.29e-7 (3.38e-7 6.65e-7)

As.=Asian, Af. =African, His= Hispanic, Eur.= European, Am. = American, f. = female, and m. = male.

LCL=lower 95% confidence limit, UCL=upper 95% confidence limit

[ns] = not significantly different from zero based on the bootstrapped confidence limits

Table 6. Hierarchal Analysis of the 2nd order terms of the mortality polynomial and heterogeneity of mortality patterns. Standard log likelihood ratio test (Chi. Sq.)

Null hypothesis Population	$c_1 = 0.0$ (3.84 needed to reject)	$c_2 = 0.0$ (3.84 needed to reject)	$c_1 = c_2 = 0.0$ (5.99 needed to reject)	$a_1 = a_2;\ b_1 = b_2;\ c_1 = c_2$ (7.81 needed to reject)
As. Am. f.	6.34	0.41[ns]	2.06[ns]	2.52[ns]
As Am. m.	0.97[ns]	0.55[ns]	1.77[ns]	0.55[ns]
His. Af. Am. f.	0.20[ns]	5.66	11.74	8.75
His.Af. Am. m.	9.52	1.30[ns]	29.41	10.74
Af. Am. f.	25.75	22.42	115.70	55.65
Af. Am. m.	43.19	3.33[ns]	66.65	58.88
His. Am. f.	24.74	21.48	86.90	21.54
His. Am. m.	25.43	3.84[ns]	81.96	82.04
Eur Am. f.	85.01	13.69	177.85	99.10
Eur Am. f. 1985	3.29[ns]	4.73	26.96	3.25[ns]
Eur. Am. f. 1986	24.93	1.04[ns]	44.54	26.65
Eur. Am. f. 1987	17.28	3.41[ns]	28.96	31.92
Eur. Am f. 1988	39.97	5.35	78.89	45.25
Eur. Am. m.	77.71	20.03	208.37	119.64
Eur Am. m. 1985	27.29	2.78[ns]	53.27	17.26
Eur. Am. m. 1986	14.95	2.52[ns]	27.39	35.63
Eur. Am. m. 1987	16.80	6.39	69.02	32.44
Eur. Am. m. 1988	15.89	10.19	57.75	37.96

As.=Asian, Af. =African, His. = Hispanic, Eur.= European, Am. = American, f. = female, and m. = male.

[ns]= not significant

Table 7. Some Characteristics of Fits of the Observed Birth Cohorts

	$1-\pi$	$\mu_2-\mu_1$	Secondary CDR	Primary CDR	Cros
As. Am. f.	0.10	170	17.2	2.5	1
As Am. m.	0.10	153	10.4	3.0	0
His. Af. Am. f.	0.12	589	52.0	3.6	1
His.Af. Am. m.	0.09	694	66.3	6.6	0
Af. Am. f.	0.13	658	59.3	6.5	0
Af. Am. m.	0.12	750	72.4	7.2	0
His. Am. f.	0.07	617	44.5	4.0	0
His. Am. m.	0.10	523	35.7	4.8	0
Eur Am. f.	0.06	487	36.7	3.0	0
Eur Am. f. 1985	0.06	488	39.0	3.1	0
Eur. Am. f. 1986	0.06	457	35.2	2.9	0
Eur. Am. f. 1987	0.06	540	39.4	2.5	0
Eur. Am f 1988	0.06	468	33.1	3.3	0
Eur. Am. m.	0.07	555	43.1	3.6	0
Eur Am. m. 1985	0.07	496	38.2	4.0	0
Eur Am. m. 1986	0.07	540	42.3	3.6	0
Eur. Am. m. 1987	0.06	639	50.4	3.3	0
Eur. Am. m. 1988	0.07	552	44.8	3.5	0

As.=Asian, Af. =African, His. = Hispanic, Eur.= European, Am. = American, f. = female, and m. = male.

CDR = subpopulation specific crude death rates per 1000 births.

Cross= whether the subpopulation and birth weight specific mortality curves cross.

Area=area between the subpopulation and birth-weight specific mortality curves as described in footnote 1.

Table 8. Ratio of Hessian to Bootstrapped Confidence Intervals

	Af. Am. f.	Eur. Am. m.
π	1.0	1.0
μ_1	1.2	1.2
μ_2	1.0	0.9
σ_1	0.9	1.1
σ_2	1.2	1.0
a_1	0.8	0.7
b_1	0.9	0.7
c_1	0.9	0.8
a_2	1.0	0.3
b_2	1.0	0.4
c_2	0.8	0.8

Af.=African, Eur. = European, Am.= American, f.= females, m.= males

Table 9. Percent Mean Bias of Parameter Estimation and Significance of Bias by Sample Size for the Three Case Studies

	Af. Am. f.			Eur. Am. m.			Simulated Parameter Sets		
	25k	50k	100k	25k	50k	100k	25k	50k	100k
π	0.20	0.13	-0.08	0.14	0.11	-0.04	0.02	0.24	-0.15
μ_1	0.00	-0.00	-0.00	0.00	-0.00	-0.00	-0.00	-0.00	-0.00
μ_2	0.02	0.01	-0.03	-0.08	-0.03	-0.02	-0.03	-0.02	-0.09
σ_1	-0.05	-0.01	0.01	-0.02	-0.02	0.01	0.01	-0.02	0.02
σ_2	-0.05	0.05	0.07	-0.08	0.03	-0.02	0.09	-0.06	0.04
a_1	-2.77	-1.68	-1.07	-2.34	0.51	0.52	-0.54	-0.08	-0.46
b_1	-2.17	-1.25	-0.85	-2.13	0.16	0.24	-0.60	-0.13	-0.46
c_1	-2.74	-1.49	-1.07	-2.96	-0.12	0.09	-0.84	-0.23	-0.58
a_2	0.79	0.38	0.24	-7.80	-6.02	-3.80	-2.45	-2.65	-0.91
b_2	0.20	0.26	0.22	-16.74	-12.44	-6.99	-6.42	-5.38	-2.27
c_2	-3.34	-0.35	0.15	-232.82	-112.97	-41.28	-51.19	-34.32	-13.55

k = 1000s, Af. = African, Eur. = European, Am. = American, f. = females, m. = males
The largest value of t for any cell is 0.63. None of the bias
estimates approach significance.

Table 10, Relative Reduction in Mean Squared Error (MSQE) with Increasing Sample Size[&] for the Three Case Studies. Values > 1.0 Indicate Improvement in Consistency as Sample Size Increases.

	Af. Am. f.		Eur. Am. m.		Simulated Parameter Sets	
	25k/50k[&]	50k/100k[&]	25k/50k[&]	50k/100k[&]	25k/50k[&]	50k/100k[&]
π	2.31	1.77	1.88	2.15	0.93	1.05
μ_1	2.15	2.00	2.01	1.94	1.08	1.01
μ_2	2.15	1.89	1.91	2.18	1.10	0.98
σ_1	2.14	1.93	2.11	1.88	1.05	1.06
σ_2	2.23	1.80	1.82	2.17	1.41	1.36
a_1	2.07	2.01	2.36	1.93	1.38	1.30
b_1	2.05	2.01	2.36	1.97	1.49	1.36
c_1	2.04	2.03	2.37	2.01	1.66	1.22
a_2	2.08	2.03	1.77	2.18	1.60	1.40
b_2	2.42	2.28	1.62	2.32	1.52	1.82
c_2	5.12	5.03	1.55	2.51	1.44	2.62
Mean	2.43	2.25	1.97	2.11	1.33	1.37

k = 1000s, Af.=African, Eur. = European, Am.= American, f.= females, m= males
&= MSQE with smaller sample size / MSQE with larger sample size, e.g. 25k/50k.

Table 11 Number of Flips in Two Case Studies per 1000 Trials

Sample size	Eur. Am. m.		Simulated Parameter Sets	
	Primary	Secondary	Primary	Secondary
25,000	2	22	11	6
50,000	0	3	5	1
100,000	0	0	2	0

Eur. = European, Am.= American, m. = males

Table 12. The Logistic Regression Coefficients for Primary Flips

Parameter	Coefficient	Std. Error	t value
Intercept	-15.40	5.20	-3.0
$1-\pi$	6.195	2.623	2.4
μ_1	4.58e-3	5.30e-4	8.7
μ_2	1.03e-3	3.95e-4	2.6
σ_1	-1.33e-2	6.60e-3	-2.0
σ_2	-5.68e-3	2.41e-3	-2.4
a_1	-1.69	0.10	-17.3
b_1	-5192.20	269.58	-19.3
c_1	-16492159.95	936477.57	-17.6
a_2	1.34	0.13	10.1
b_2	3044.59	180.97	16.8
c_2	8981036.11	601720.63	14.9

Null Deviance: 2174.0 on 999 degrees of freedom
Residual Deviance: 332.2 on 988 degrees of freedom

Table 13. The regression coefficients for secondary flips

Need to rerun to get more significant places (Mike had the results rounded)

Parameter	Coefficient	Std. Error	t value
Intercept	-11.82	5.01	-2.4
$1-\pi$	-37.28	3.27	-11.4
μ_1	-1.90e-3	6.09e-4	-3.1
μ_2	4.88e-3	4.87e-4	10.0
σ_1	8.19e-3	7.41e-3	1.1
σ_2	-3.65e-3	2.36e-3	-1.5
a_1	9.61e-2	4.35e-2	2.2
b_1	158.85	89.77	1.8
c_1	1199184.55	469328.61	2.6
a_2	-7.21e-1	1.50e-1	-4.8
b_2	-1543.91	137.87	-11.2
c_2	-7095120.50	535387.72	-13.3

Null Deviance: 1134.2 on 999 degrees of freedom
Residual Deviance: 545.3 on 988 degrees of freedom

Table 14 – Power Estimates for Two Case Studies

Sample Size	12,500	25,000	50,000	100,000
Af. Am. f.				
Power	82.5	98.1	100.0	100.0
95% Confidence Limits	80.0 to 84.8	97.1 to 98.9	99.6 to 100.0	99.6 to 100.0
Eur. Am. m.				
Power	21.8	47.3	79.3	97.0
95% Confidence Limits	19.3 to 24.5	44.2 to 50.5	76.7 to 81.8	95.7 to 98.0

Results based on 1000 trials.

Table 15 – Power Estimates for the Remaining 10 Independent Observed Data Sets

	12,5000 (95% CL)	25,000 (95% CL)	50,000 (95% CL)	100,000 (95% CL)
As. F.	58	84	98	100
	(43.2 to 71.8)	(70.9 to 92.8)	(89.4 to 100)	(92.9 to 100)
As. M.	28	40	72	98
	(16.2 to 42.5)	(26.4 to 54.8)	(57.5 to 83.8)	(89.4 to 100)
Af. Am. m.	72	100	100	100
	(57.5 to 83.8)	(92.9 to 100)	(92.9 to 100)	(92.9 to 100)
Af His.c m.s	90	100	100	100
	(78.2 to 96.7)	(92.9 to 100)	(92.9 to 100)	(92.9 to 100)
Eur. His. f.	44	70	100	100
	(30.0 to 58.8)	(55.4 to 82.1)	(92.9 to 100)	(92.9 to 100)
Eur. His. m.	38	70	96	100
	(24.7 to 52.8)	(55.4 to 82.1)	(86.3 to 99.5)	(92.9 to 100)
Eur Am. f. 1985	14	36	44	96
	(5.8 to 26.7)	(22.9 to 50.8)	(30.0 to 58.8)	(86.3 to 99.5)
Eur. Am. f 1986	32	76	94	100
	(19.5 to 46.7)	(61.8 to 86.9)	(83.5 to 98.7)	(92.9 to 100)
Eur. Am. f. 1987	20	48	88	100
	(10.0 to 33.7)	(33.7 to 62.6)	(75.7 to 95.5)	(92.9 to 100)
Eur. Am. f. 1988	58	88	100	100
	(43.2 to 71.8)	(75.7 to 95.5)	(92.9 to 100)	(92.9 to 100)

Estimates based on 50 trials each

As.=Asian to Af. =African, His. = Hispanic, Eur.= European, Am. = American, f. = female, and m. = male.

Table 16 – Logistic Regression of Power based on Data in Tables 7 and 15

.

Parameters	Coefficients	Standard Error	t-value
Intercept	-10.4	0.8	-12.7
Difference in Areas[1]	5.2e-6	5.7e-7	9.1
Primary Crude Death Rate	487	72	6.8
Difference in birth weight means	3.6e-3	6.1e-4	5.8
Indicator: crossing mortality curves	.74	.30	2.5
Sample Size	7.4e-5	4.8e-6	15.4

Null Deviance: 914.7 on 39 df
Residual Deviance: 84.2 on 34 df

Table 17 Comparison of Percent Power Predictions from Logistic Regression (Table 16) and Direct Estimation (1000 simulated trials) for the Two Case Studies (Table 14).
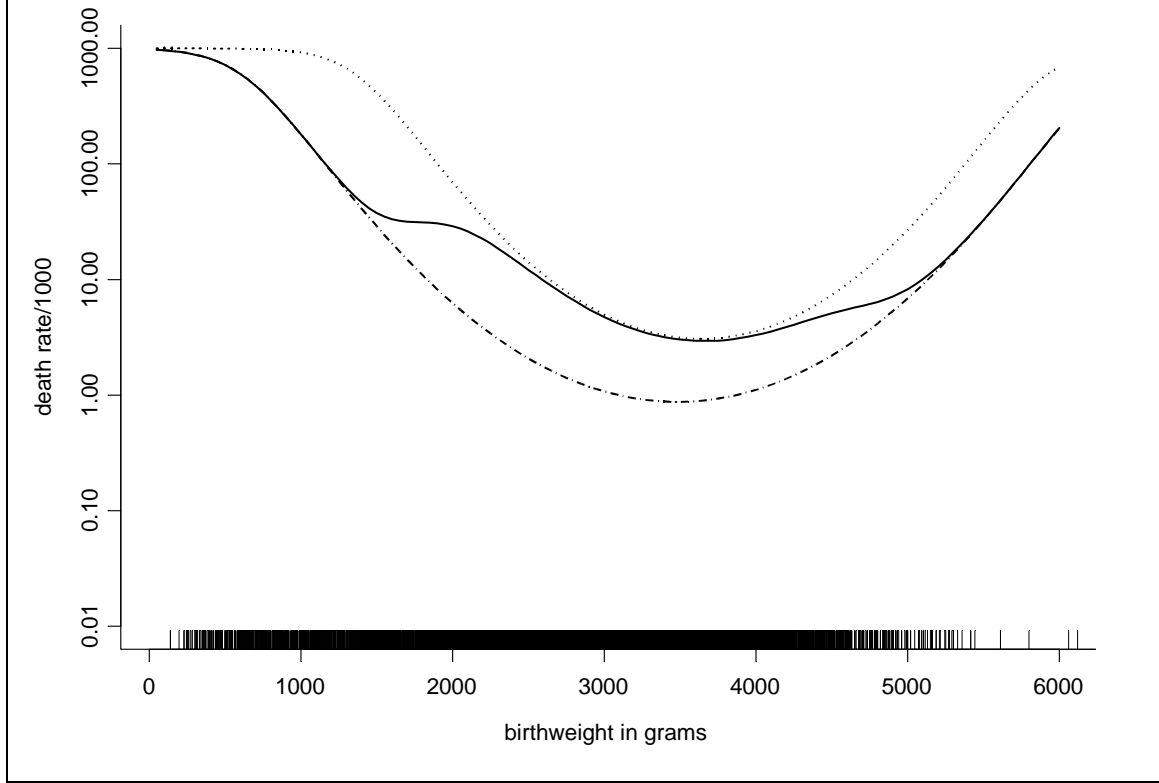
Sample Size	From Regression (95% Confidence Limits)	From 1000 Simulations (95% Confidence Limits)
	Af. Am. f.	
12,500	71.8	82.5
	(64.3 to 79.3)	(80.0 to 84.8)
25,000	86.5	98.1
	(82.2 to 90.8)	(97.1 to 98.9)
50,000	97.6	100
	(96.5 to 98.7)	(99.6 to 100)
100,000	99.9	100
	(99.9 to 100.0)	(99.6 to 100)
	Eur. Am. m.	
12,500	17.3	21.8
	(12.6 to 22.0)	(19.3 to 24.5)
25,000	34.5	47.3
	(28.0 to 41.0)	(44.2 to 50.5)
50,000	76.9	79.3
	(70.9 to 82.8)	(76.7 to 81.8)
100,000	99.3	97.0
	(98.7 to 99.8)	(95.7 to 98.0)

Af. = African, Eur. = European, f. = female, m. = male.

African Females ('85-'88)

probability density

birthweight in grams
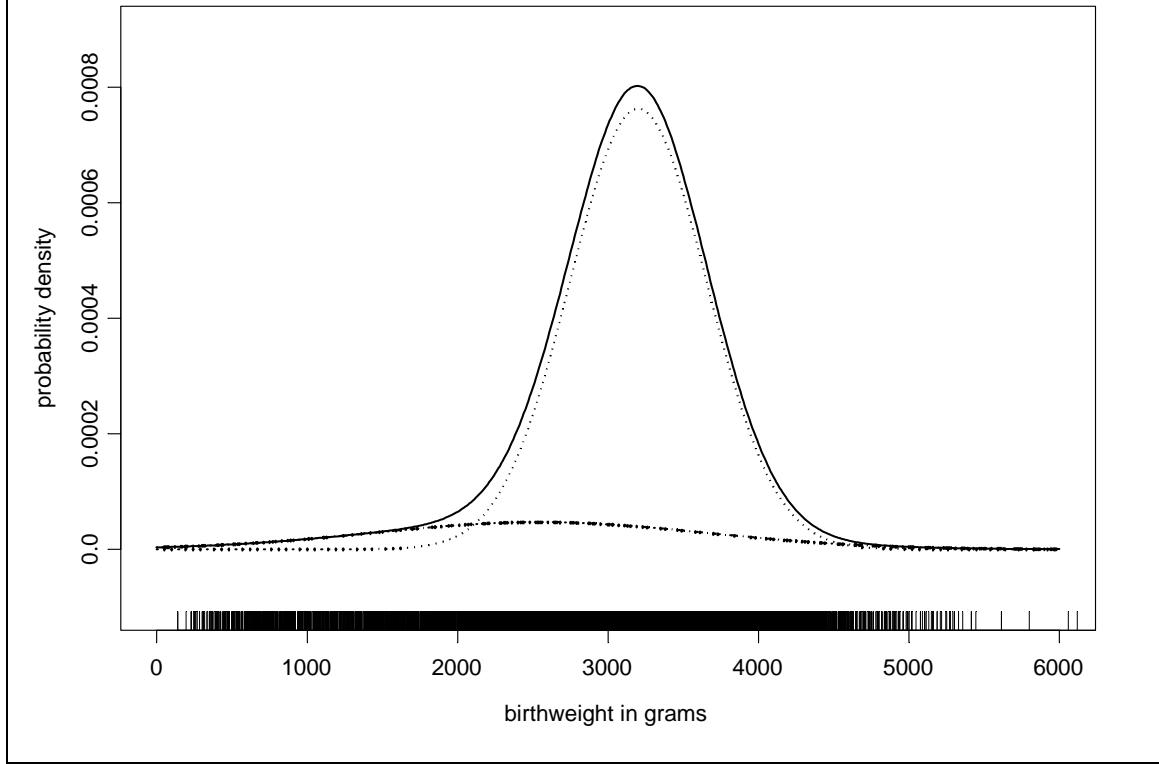
African Females ('85-'88)
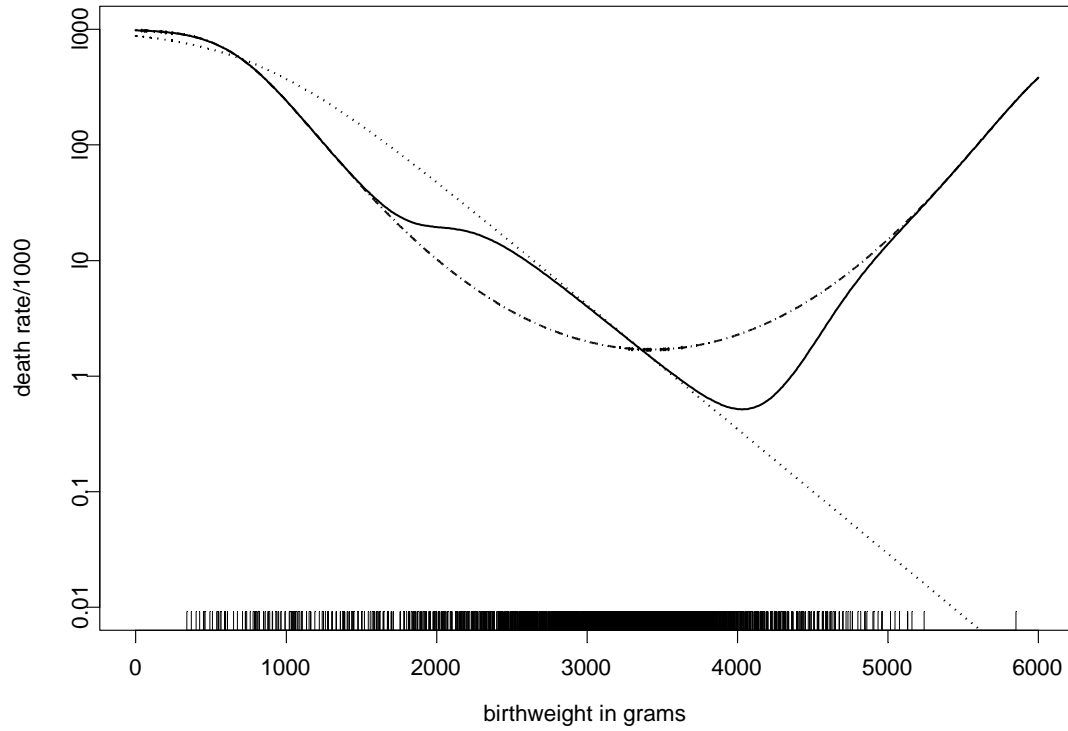
African-Hispanic Females ('85-'88)

death rate/1000

birthweight in grams
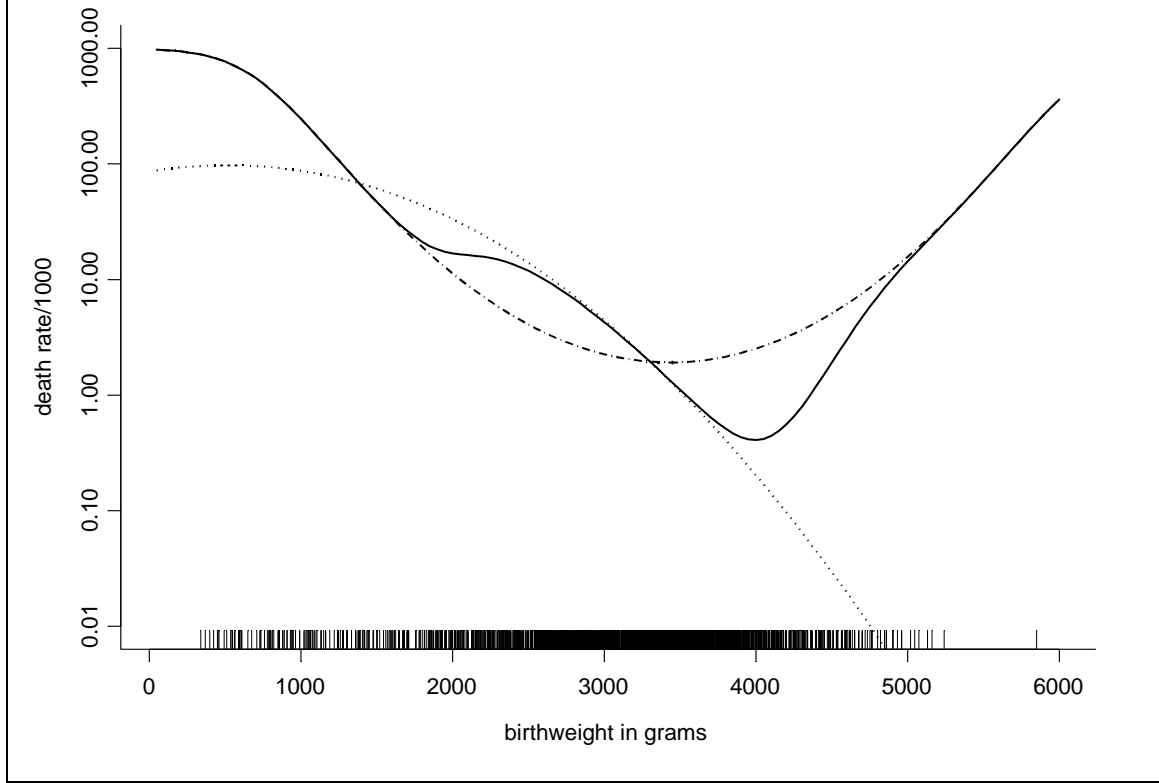
African-Hispanic Females ('85-'88)

African Females ('85-'88)