Send correspondence to:

Timothy B. Gage

Department of Anthropology, AS 114

University at Albany – SUNY

Albany, NY 12222

Office:518-442-4704

Fax: 518-442-5710

Email: tbg97@albany.edu

Education, birth weight, and infant mortality

Timothy B. Gage [1,2,3]

E. O'Neill [1]

Fu Fang [1]

Howard Stratton [2]

Institutional affiliations: 1) Department of Anthropology, University at Albany-SUNY, 2) Department of Epidemiology and Biostatistics, University at Albany-SUNY. 3) Department of Genetics, Southwest Foundation for Biomedical Research.

Abstract

This research determines if the declines in infant mortality with increased maternal education are due to "indirect" effects that operate through improved birth weight or to birth weight independent "direct" effects. The data used are the US national linked birth death files, African, Mexican and European American cohorts, 2001. Education is dichotomized as completing 12th grade or less. The analysis is conducted using Covariate Density Defined mixture of logistic regressions, which considers "normal" and "compromised" births separately. Among "normal" births, mean birth weight increases significantly with education by 27-108 grams. Mortality declines due to "direct" effects of education (log-odds -0.04 to -0.91). The "indirect" effects generally increase infant mortality (log-odds -0.17 to 0.45) despite improvements in birth weight. No consistent birth weight or mortality effects occur among "compromised" births. These results are consistent with the view that birth weight is not on the causal pathway to mortality.

Introduction

Measures of socioeconomic status, such as education are thought to be important correlates of body size, e.g. stature and mortality (Floud, Wachter and Gregory 1990). What is not clear is whether the effects of socioeconomic status influence mortality "directly" or "indirectly" through its influence on the anthropometric measures. A well-know example is the relationship of socioeconomic level with birth weight and infant mortality. The correlation between birth weight and infant mortality is empirically so well established that the current US policy to reduce infant mortality is to improve birth weight (Buehler et al. 1987; Institute-of-Medicine 1985; Mc Cormick 1985; US-DOHHS 2000). However, many theoreticians have argued that birth weight is not on the causal pathway to infant mortality (Mosely and Chen 1984; Wilcox and Russell 1990; Wise 2003) at least among "normal" births.

Recently a statistical method of determining if birth weight is or is not on the "causal" pathway to infant mortality has been developed (Gage et al. 2004), based on the Wilcox (Wilcox and Russell 1990) definition of "causality". Wilcox argues that among "normal" births when the birth weight distribution shifts due to a stressor (e.g. smoking during pregnancy; fetal development at altitude) that the birth weight specific mortality curve shifts in the same direction a similar amount so that there is no net change in infant mortality due to the shift in the birth weight distribution (Figure 1a). Consequently, birth weight is not on the "causal" pathway to infant mortality. In addition, he argues that there may be a consistent increase (or decline) in infant

mortality at all birth weights (independent of birth weight) due to a stressor (e.g. smoking during pregnancy; but not fetal development at altitude) (Figure 1b). Based on this definition it only need be shown that the shift in the birth weight distribution among "normal births" is not matched by a similar shift in the birth weight specific infant mortality curve to demonstrate that birth weight is on the "causal" pathway to infant mortality (Figure 1c). This will be referred to here as an "indirect" effect through birth weight of the stressor on infant mortality. The birth weight independent effect will be called a "direct" effect of the stressor. Wilcox does not discuss whether "compromised" births behave in the same manner as "normal" births. Nevertheless to fully examine Wilcox's hypothesis requires a method that distinguishes between "normal" and "compromised" births. Covariate Density Defined mixture of logistic regressions (CDDmlr) distinguishes between "normal" and "compromised" births and can estimate the significance of "indirect" as well as "direct" effects.

Figure 1 about here

The aim of this research is to determine if education, a measure of socioeconomic level, affects infant mortality "directly", or "indirectly" though birth weight, or both. The analyses are conducted on three populations by sex, African, European, and Mexican Americans from the 2001 US birth cohort.

Data and Methods

The data for these analyses was obtained from the national linked birth death files for the birth cohort born in 2001. Births with missing values, that is missing birth

weights, maternal education, or race/ethnic designations, are excluded. Race and ethnic origin is based on mother's reported race and ethnic origin. Summary statistics for the six 2001 birth cohorts are presented in Table 1. Confidence intervals are estimated using bootstrap procedures. Two hundred bootstrap samples of approximately 200,000 births are selected for each population by sex from the complete 2001 birth cohort to estimate standard errors of the estimates. In each case the proportion of births from the full cohort necessary to achieve a slightly larger than 200,000 sample is estimated and then each birth is randomly included (or excluded) based on a random uniform number between 0.0 and 1.0. Samples of 200,000 are used instead of the conventional procedure of sampling with replacement samples the size of the original cohort due to the large computational costs of repeatedly fitting such large cohorts.

Table 1 about here

The analysis is conducted using Covariate Density Defined mixture of logistic regressions (Gage et al. 2004). Conceptually, this method fits a multi-component Gaussian mixture model to the birth weight distribution, which divides the population into several (2 in this case) latent subpopulations, and simultaneously fits a separate logistic regression to each latent subpopulation. This procedure is useful when the Gaussian mixture model accounts for unobserved heterogeneity. In the case presented here, birth weight is the covariate to which a two-component Gaussian mixture model is fitted. The subpopulation accounting for the majority of individuals is labeled the primary subpopulation (Figure 2a) and is considered to represent births undergoing "normal" fetal development (Gage et al. 2004). The remaining subpopulation accounts

for most low birth weight and macrosomic births (Figure 2a) and is consequently interpreted as accounting for births undergoing "compromised" fetal development (Gage et al. 2004). The logistic regressions on infant mortality for each subpopulation are generally parameterized as second-degree polynomials of birth weight to account for the reverse J-shaped relationship of birth weight and infant mortality (Figure 2b). Applications indicate that the two components identified by the Gaussian mixture model are heterogeneous with respect to infant mortality (Gage et al. 2004). The "compromised" population consistently displays lower birth weight specific mortality (Figure 2b), but higher overall mortality due to the less favorable distribution of birth weight among "compromised" births (Gage et al. 2004).

Figure 2 about here

Here we expand this basic model by adding exogenous covariates to the Gaussian mixture model and to the logistic regressions. In the case presented here the covariate is education: high school education or less (coded 0) versus higher education (coded 1). In the Gaussian mixture model, all parameters, i.e. the mixing proportion, means and variances, are each defined as a function of education. In addition, education is added to the logistic regression models, as a covariate on the constant, linear and squared terms of a second-degree polynomial of birth weight. Finally we define birth weight within each logistic regression as the Z-score based upon the corresponding Gaussian subpopulation that it represents. The result is a model that can test Wilcox (Wilcox and Russell 1990) definition of "direct" and "indirect" effects (Figure 1). After standard transformation of the quadratic covariates to account for correlation inherent

in this specification, the education interaction covariate on the constant of the birth weight polynomial can be interpreted as a direct effect (constant at all birth weights). The education interaction covariate on the linear birth weight term accounts for the shift, left or right, in the birth weight specific mortality curve with respect to mean birth weight. Due to the use of Z-scores of birth weight, an insignificant interaction on the linear birth weight interaction term indicates that the birth weight specific mortality curve shifts in concert with birth weight density, i.e. no indirect effect of education as argued by Wilcox. The education interaction on the squared term accounts for changes in shape of the birth weight specific infant mortality curve relative to the birth weight distribution. This is also an indirect effect but is not considered in Wilcox's original theory. However, it is related to Basso et al's (Basso, Wilcox and Weinberg 2006) extension of Wilcox's theory, which attributes this effect to confounding rather than causality. Again an insignificant interaction term indicates no indirect (or confounding) effect, due to the use of Z-scored birth weight. CDD mixture of logistic regression can be used to explore Basso's confounding hypothesis. Here, however, we examine shift and shape affects together, as potential "causal" effects. A formal definition of this model is presented below.

The probability of death (y) is a product of a) the distribution of the birth weight (x) given the exogenous dichotomous covariate (z), and b) the conditional mortality given x and z:

$$f(y,x,z;\beta,\theta) = f_2(y \mid x,z;\beta,\theta) \times f_1(x \mid z;\theta) \qquad \text{(Eq. 1)}$$

In the case of two Gaussian subpopulations (labeled as p and s), the $f_1(x \mid z;\theta)$ is expressed as:

$$f_1(x \mid z;\theta) = f_1(x \mid z;\pi_s,\mu_s(z),\sigma_s^2(z),\mu_p(z),\sigma_p^2(z))$$

$$= \pi_s(z) \times N_{500}(x;\mu_s(z),\sigma_s^2(z)) + (1-\pi_s(z)) \times N_{500}(x;\mu_p(z),\sigma_p^2(z)) \quad \text{(Eq. 2)}$$

$$\pi_s(z) = \pi_{s0} + z \times \pi_{s1} \quad \text{(Eq. 3)}$$

$$\mu_i(z) = \mu_{i0} + z \times \mu_{i1} \quad \text{(Eq. 4)}$$

$$\sigma_i(z) = \sigma_{i0} + z \times \sigma_{i1} \quad \text{(Eq. 5)}$$

where π_s is the proportion of births belonging to the subpopulation s. For $i = p$ and s, N_{500} represents the Gaussian density, truncated at 500 grams, with mean μ_i and variance σ_i. The probability of death conditioned on x and z is given by:

$$f_2(y=1 \mid x,z;\beta,\theta) = f_2(y=1 \mid x,z;\beta_p,\beta_s,\theta)$$

$$= q(x \mid z,\theta) \times P_p(x \mid z,\beta_p) + (1-q(x \mid z,\theta)) \times P_s(x \mid z,\beta_s)$$

$$= q(x \mid z,\theta) \times P_p(x_p^* \mid z,\beta_p^*) + (1-q(x \mid z,\theta)) \times P_s(x_s^* \mid z,\beta_s^*) \quad \text{(Eq. 6)}$$

where x_i^* is the standardized x using μ_i and σ_i, P_i is the probability of death for an infant with birth weight (x or x_i^*) and covariate z in the subpopulation i given by a quadratic logistic form:

$$P_i(x \mid z,\beta_i) = P_i(x \mid z,a_i,b_i,c_i) = \frac{exp(a_i(z) + x \cdot b_i(z) + x^2 \cdot c_i(z))}{1 + exp(a_i(z) + x \cdot b_i(z) + x^2 \cdot c_i(z))} \quad \text{(Eq. 7)}$$

$$a_i(z) = a_{i0} + z \times a_{i1} \quad \text{(Eq. 8)}$$

$$b_i(z) = b_{i0} + z \times b_{i1} \quad \text{(Eq. 9)}$$

$$c_i(z) = c_{i0} + z \times c_{i1} \quad \text{(Eq. 10)}$$

q is the conditional probability of that infant belonging to the subpopulation s. The mixture submodel $f_1(x \mid z;\theta)$ (Eq. 2) determines that

$$q(x \mid z;\theta) = \frac{\pi_s(z) \times N_{500}(x;\mu_s(z),\sigma_s^2(z))}{f_1(x \mid z;\theta)} \quad \text{(Eq. 11)}$$

The function $f(y,x,z;\beta,\theta)$ (Eq. 1) was fit to individual data by the method of maximum likelihood. Overall, there are 22 parameters for the model with a dichotomous covariate z and only 11 for the model without z. Bias-adjusted 95% CIs for the parameter, and in some cases combinations of parameters, are estimated using two bootstraps of 100 repetitions each. Due to the relatively small number of repetitions in the bootstraps and potential for instability in marginally significant and marginally insignificant results borderline results are acknowledged in the text.

Table 2 about here

Results

Education influences the birth weight distribution thorough changes in the mean and standard deviation of both primary and secondary birth weight (Figure 3, Table 3). The mean of the primary subpopulation increases significantly with education in all birth cohorts examined. The increase is largest in European Americans and smallest in Mexican Americans. The standard deviation of the primary subpopulation changes significantly in all birth cohorts. This declines in all cohorts examined except African American males. The mean of the secondary subpopulation declines in African

American birth cohorts about 100 grams, while it increases in the European American birth cohorts by more than 200 grams. The secondary mean does not change significantly among the Mexican American birth cohorts. Finally, the standard deviation of the secondary subpopulations changes significantly in all populations except Mexican American females. Among European Americans and Mexican Americans it declines. However, among African American birth cohorts it increases with education. Finally, education is associated with an increase the proportion of primary births, although this is typically less than 1%. Nevertheless the increase is significant in three out of the six populations examined (i.e. African American males and females, and European American females). Note a) that the response of African American birth weight distributions to education differs qualitatively particularly in the secondary subpopulation and mixing proportion from the responses of European American birth cohorts and b) that the response of the Mexican American birth weight distributions to education is smaller than the responses of the other ethnic birth cohorts.

Figure 3 about here

Table 3 about here

The logistic regression results indicate that mortality declines due to higher education in both primary and secondary subpopulations. However, most of the effect is due to primary births. The results are shown as log odds ratio in Table 4. In particular, in 5 of the 6 populations examined, total primary mortality declines significantly with education. The effect is smallest among the primary Mexican American birth cohorts in general. On the other hand, the decline in total secondary

mortality with higher education is only significant in 2 of the 6 subpopulations, Mexican American females and European American males. To assess the relative contribution of primary and secondary births to the total decline in infant mortality due to education, the log odds ratios in Table 4 must take into account that the primary subpopulation includes >90% of all births (relative contribution Table 4). Comparison of these adjusted estimates indicates that the effect of higher education occurs large in the primary subpopulation even among Mexican American births where the effects of education are smallest.

Table 4 about here

In the primary subpopulation, the main influence of education on infant mortality is due to direct effects. The direct effects of education are all negative, that is higher education tends to reduce mortality (Table 4). These are significant in all birth cohorts examined except European American and Mexican American males. The indirect effects (i.e. the combined effects of shift and shape) are predominately positive (i.e. to increase mortality), but are only significant in three cases, Mexican American males and females, and African American females. Note that the result for African American males is marginally insignificant. In two cases Mexican American males and European American males, the effects of education tend to reduce mortality. The decline in Mexican American males is significant, but only marginally, while the decline in European American males is not significant. In most populations, the birth weight specific indirect effects of education are concave, that is mortality increases at lower and higher birth weights (Figure 4a). The direct effect is of course independent of birth

weight as indicated by a flat line in Figure 4a and 4b. Among Mexican American cohorts, however, the birth weight specific indirect effects of education are convex, that is the indirect effect on mortality is highest at normal birth weights and lower at low and high birth weights (Figure 4b). The difference between Mexican American males and females is that the curve is slightly higher among females so that the indirect effects exceed a log-odds of 0.0 in the normal birth weight range, where as the male curve does not. All of the primary indirect effects appear to be due to changes in the shape of the mortality curve relative to the density of birth weight as opposed to a horizontal shift in the mortality curve relative to the density of birth weight which should produce a monotonically increasing or decreasing log odds ratio with respect to birth weight.

Figure 4 about here

Decomposition of secondary births is similar in that the direct effects of higher education reduce mortality, while the indirect effects of secondary mortality tend to increase mortality, but the results are not generally significant. The direct effects are only significant in three populations, while the indirect effects are only significant in two populations (Table 4). Again, the direct effects tend to be larger than the indirect effects. The birth weight specific trends in indirect effects are not consistent across populations.

Finally, the impact of education on the mixing proportion (a cohort "composition" effect) also influences total mortality, but only slightly. The proportion of primary births increases with education significantly in African American males and females and European American females. Since the primary subpopulation has lower

overall mortality, this effect explains some of the decline in total mortality due to education. All together this phenomenon accounts for about 10.6% of the total decline in infant mortality due to higher education. The effects are greatest in African American males where it accounts 27.4% of the decline. Overall however, these effects are small.

Discussion and Conclusions

A theoretical limitation of the analysis presented above is that Wilcox's original theory does not completely account for all of the potential influence of birth weight on infant mortality. In particular, Wilcox's original theory assumes that the reverse J-shaped birth weight specific mortality curve is constant. It is possible that birth weight is causally responsible for the reverse J-shape. Recently, Basso (Basso et al. 2006) has provided an extension of the original theory that attributes the reverse J-shape to confounding. CDD mixtures of logistic regression can be used to explore this possibility as well. However, in the present analysis we test Wilcox's original theory, although the model used allows the shape of the reverse J-shape to change in response to the covariate education. We have interpreted this change in shape as a kind of indirect effect since it involves birth weight and as such included it as potentially due to a "causal" effect of birth weight along with shift effects (causality) in Wilcox's original theory. It should be noted that the important "indirect" effects that we identified above appear to be due to shape changes and not shift changes.

A methodological limitation of the analyses presented above is limiting the bootstraps to 200,000 as opposed to resampling at the level of the national populations.

Smaller bootstrap sample sizes were chosen simply to reduce the computational time required to complete the analysis. The number of bootstrap replicates was further restricted to 200 (100 to estimate the bias, and 100 to estimate the confidence regions) for the same reason. Increasing the bootstrap sample size would increase power allowing us to conclude that smaller effects are consistent, while increasing the number of replicates might make the inconsistencies in significance among the populations more consistent, particularly those where significance or insignificance is marginal. Consequently, marginal results are noted above. On the other hand, the procedure calls attention to those effects that are most important and downplays those with small effect sizes. In any event samples of 200,000 are already quite large.

The analyses presented above suggest that education has little effect on Mexican American birth cohorts compared to other birth cohorts. The effect of higher education on the birth weight distribution tends to be small and is more often insignificant. The impact of education on primary mortality is smaller than in other cohorts as well. Furthermore, the trends in primary indirect effects with birth weight are qualitatively different from other populations (Figure 4). It is not clear why Mexican American infants do not respond to the level of maternal education like other birth cohorts. However, Mexican American birth cohorts have the lowest observed infant mortality rate of any of the populations examined (Table 1). In addition, approximately 64% of Mexican American births are to Mexican-born mothers. Perhaps the differences in response to education noted above are a part of the "nativity complex" that is thought to influence infant mortality in recent migrant populations (Hummer et al. 1999).

The most important influences of higher education in infant mortality appear to be due to direct effects, that is effects that are independent of birth weight, among "normal" (primary) births. In all cases these effects tend to reduce mortality with increased educational level. The indirect effects of higher education are relatively small, less significant and tend to increase infant mortality with higher education. This is surprising since birth weight increases significantly, and the variance in birth weight declines with higher education in most primary subpopulations examined which reduces the number of low birth weight infants and is therefore expected to reduce mortality. We attribute this negative indirect effect of higher education to relaxed fetal selection on unobserved heterogeneity with higher levels of education, due perhaps to increased medical care during pregnancy.

On the other hand, "compromised" (secondary) births contribute substantially less to the decline in infant mortality with higher education. Here there are also significant effects on the mean and variance of birth weight, but in different directions in different populations. Secondary birth weight densities deteriorate with higher education, that is mean birth weight declines and variance in birth weight increases among secondary African American births, while the opposite occurs in European American births. The direct effects still tend to reduce mortality, and the indirect effects still tend to increase mortality but the effects are smaller, and less often significant. In this case the African American deterioration in birth weight is consistent with the increase in indirect mortality with higher education among "compromised" births.

Finally, higher education tends to reduce the proportion of "compromised" births, with a resulting decline in overall infant mortality at least in European and African American birth cohorts. However, these effects are only a small fraction of the total education effect.

The results presented here generally support Wilcox's (Wilcox and Russell 1990) original theory that birth weight is not on the causal pathway to infant mortality. A sufficient number of significant indirect effects among "normal" births that influence infant mortality through birth weight were found, which might falsify Wilcox's original hypothesis. In particular the significant cases are not attributable to a horizontal shift in the birth weight specific mortality curve relative to birth weight. They are more consistent with a change in the shape of the birth weight specific mortality curve relative to the birth weight density. The lack of horizontal shift is in agreement with Wilcox's (Wilcox and Russell 1990) original theory that birth weight is not on the causal pathway. The changes in the reverse-J-shape could be due to "causal" effects of birth weight not included in Wilcox's original theory or to confounding as per Basso et al's. (Basso et al. 2006) extension of Wilcox's original theory. Basso's extension could be explored further using CDD mixtures of logistic regressions. It is interesting to point out, however, that if these indirect effects are causal, then the causal effect is opposite what is generally proposed and opposite the assumption upon which our current national policy is based. The indirect effect of education through improved birth weight (higher birth weight, lower variance in birth weight) on infant mortality tends to increase infant mortality, not decrease infant mortality, net of the direct effects of

education. Additional analyses will be necessary to determine if these trends are consistent, are really "causal" or are due to confounding. In any event the results suggest that our National policy of improving birth weight in an attempt to improve infant mortality may not be effective.

Acknowledgement

This work is supported by NIH grant R01 HD037405.

References

Basso, O., A. Wilcox, and C. Weinberg. 2006. "Birth weight and mortality: Causality or confounding?" Pp. 303-311 in *Am. J. of Epidemiol.*

Buehler, J.W., J.C. Kleinman, C.J.R. Hogue, L.T. Strauss, and J.C. Smith. 1987. "Birth weight-specific infant mortality, United States, 1960 and 1980." Pp. 151-161 in *Public Health Reports*.

Floud, R., K. Wachter, and A. Gregory. 1990. "Height, Health and History." in *NBER Series on Long-term Factors in Economic Development*, edited by C.L. Pope. Cambridge: Cambridge University Press.

Gage, T.B., M.J. Bauer, N. Heffner, and H. Stratton. 2004. "The pediatric paradox: Heterogeneity in the birth cohort." Pp. 327-342 in *Human Biology*.

Hummer, R.A., M. Biegler, P.B. DeTurk, D. Forbes, W.P. Frisbie, Y. Hong, and S.G. Pullum. 1999. "Race/ethnicity, nativity, and infant mortality in the United States." Pp. 1083-1118 in *Social Forces*.

Institute-of-Medicine. 1985. "Preventing Low Birthweight." Washington: National Academy Press.

Mc Cormick, M.C. 1985. "The contribution of low birth weight to infant mortality and childhood morbidity." Pp. 82-90.

Mosely, W.H.and L.C. Chen. 1984. "An analytical framework for the study of child survival in developing countries." in *Population and Development Review Supplement to Vol. 10.*, edited by L.C. Chen. Cambridge: Cambridge University Press.

US-DOHHS. 2000. "Healthy People 2010: Understanding and improving health." Pp. 91-95. Washington, DC: US Department of Health and Human Services.

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." Pp. 7-10 in *J. of Ped.*

Wise, P.H. 2003. "The anatomy of a disparity in infant mortality." Pp. 341-362 in *Annual Review of Public Health*.

Figure Captions

Figure 1. Graphical representation of Wilcox's definition of "causality". Panel (a) represents a shift in birth weight that is accompanied with a shift in the birth weight specific mortality curve so that no change in mortality occurs (birth weight is not "causal"). Panel (b) represents a birth weight independent change in infant mortality, (direct effect, birth weight is not "causal"). Pane (c) represents a shift in birth weight that is not accompanied by an identical shift in the birth weight specific mortality curve so that mortality does change.

Figure 2. Graphical representation of the Covariate Density Defined mixture of logistic regressions as applied to birth weight. The results presented are African American males 2001 based on analyses developed below in this paper. Panel (a) represents the density of total birth weight as the sum of two Gaussian densities, primary and secondary. The secondary is considered "compromised" since it accounts for the majority of low birth weight and macrosomic infants. Panel (b) represents characteristic total, as well as, primary and secondary specific mortality curves.

Figure 3. Shift in birth weight density due to higher education: African American males 2001. African American females, and European American males and females are similar. The shift for Mexican Americans is much smaller.

Figure 4. Log odds ratios for direct, indirect, and total effects of higher education on infant mortality. Positive effects indicate increased mortality, while positive effects represent reduced mortality due to higher education. Panel (a) presents African American females. These trends are characteristic of all African and European American

populations examined. Panel (b) presents Mexican American females. Mexican American males are similar.

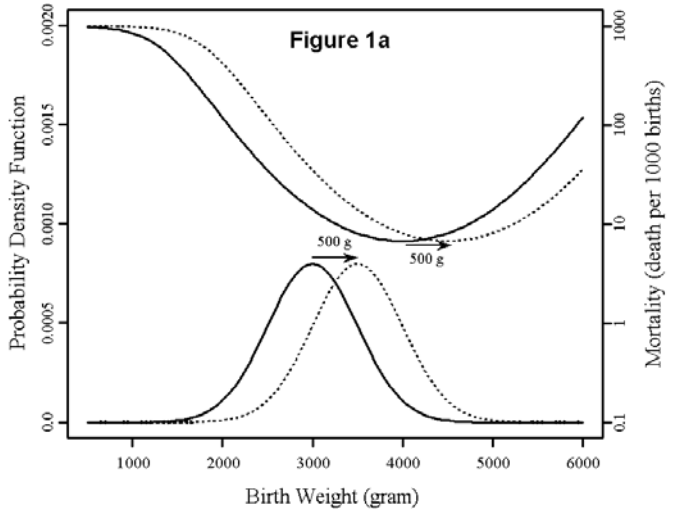
Figure 1a
Probability Density Function
Mortality (death per 1000 births)
Birth Weight (gram)
500 g
500 g

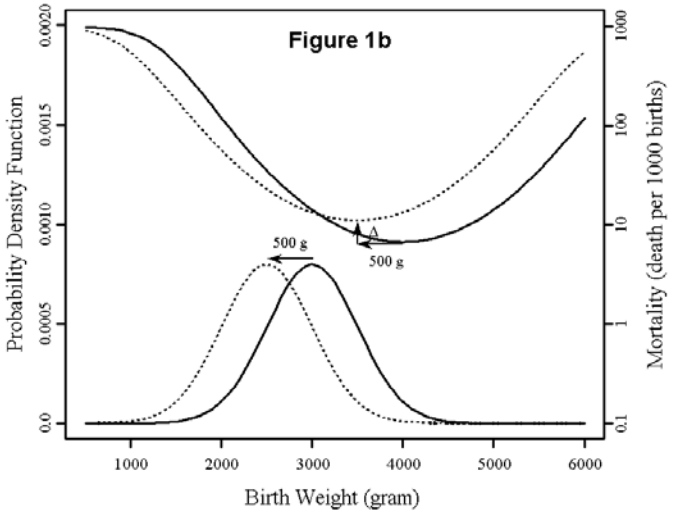
Figure 1b
Probability Density Function
Mortality (death per 1000 births)
Birth Weight (gram)
500 g
500 g

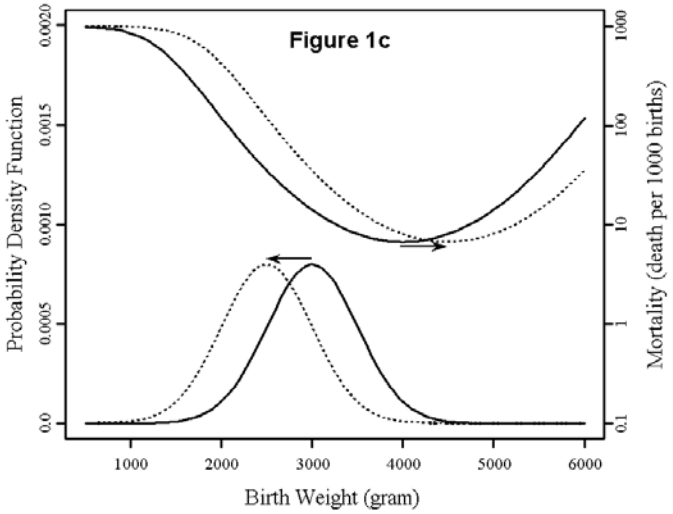
Figure 1c
Probability Density Function
Mortality (death per 1000 births)
Birth Weight (gram)

Figure 2a. African American Males: Probability Densities

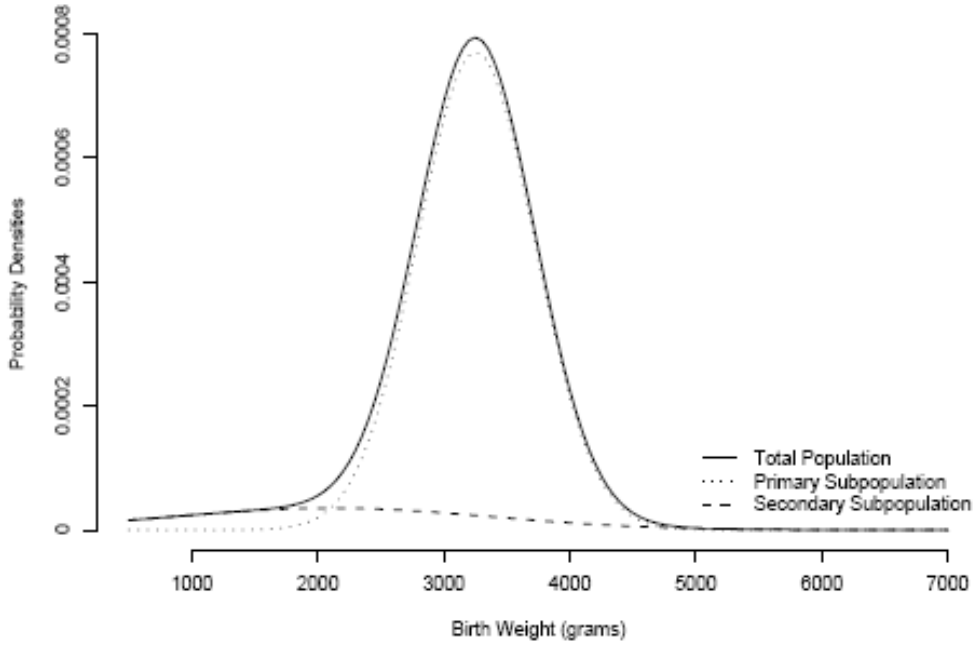

Figure 2b. African American Males: Mortality

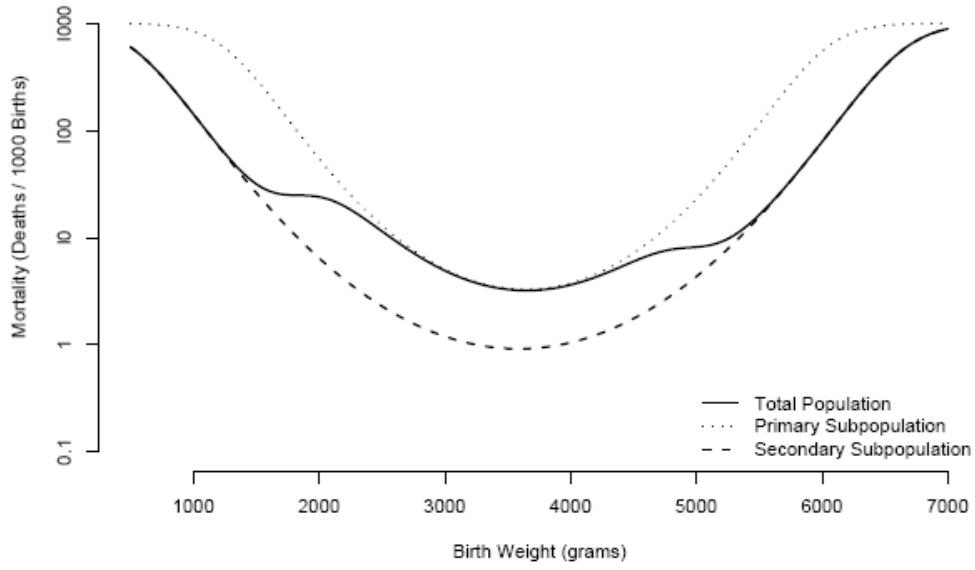

Figure 3. African American Males: Total Probability Densities

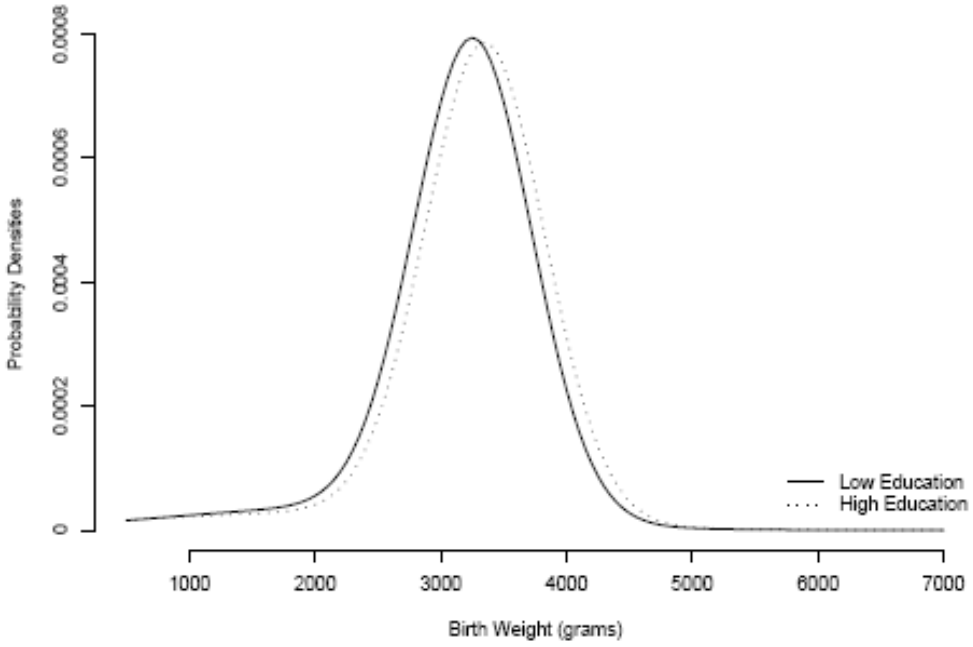

Figure 4a African American Female Primary Subpopulation Decomposition

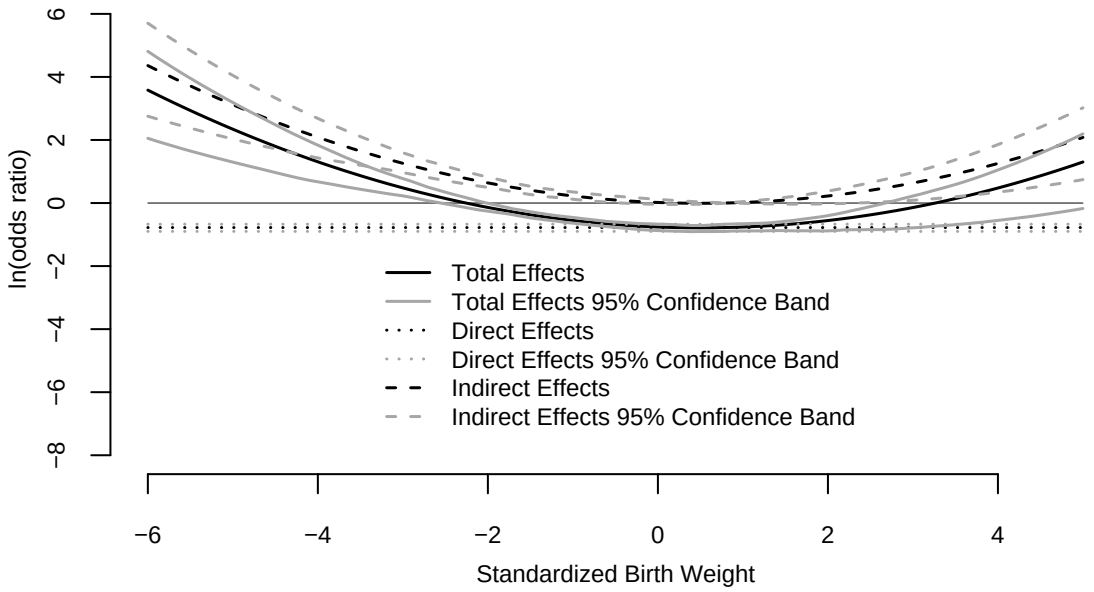

Figure 4b Mexican American Female Primary Subpopulation Decomposition

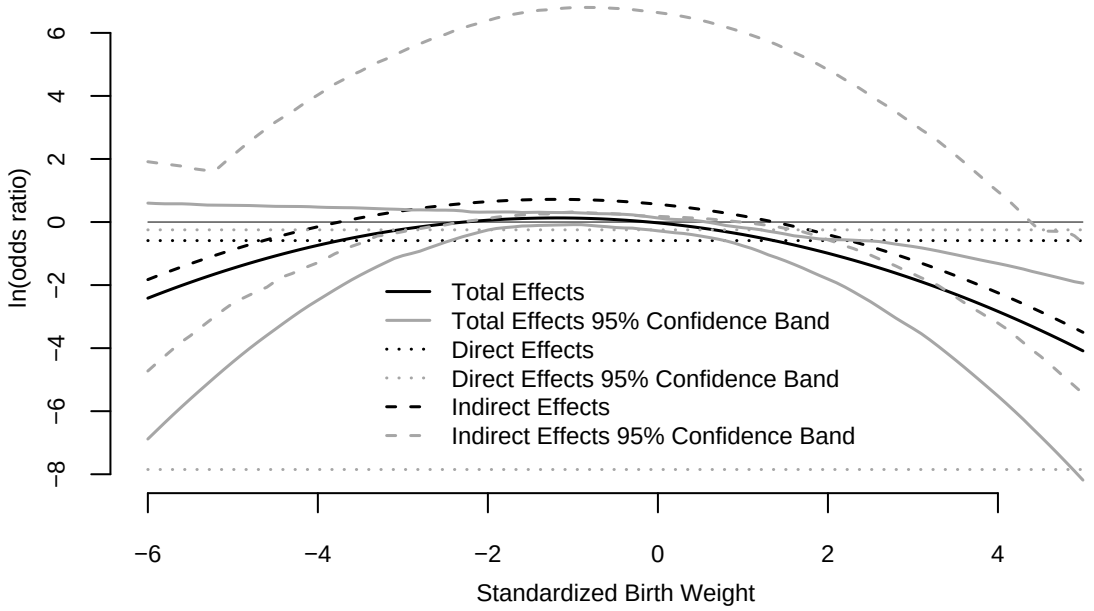

Table 1 Descriptive statistics for the sample populations

Birth Cohort	# Total Births	NA of bwt *		NA of education &		NA of bwt and education *,&		# Births Used	CDR
		# of births	% of births	# of births	% of births	# of births	% of births		
Non-His. Eur. Am. F.	1,093,096	69,513	6.36	8,507	0.78	77,173	7.06	1,015,923	3.42
Non-His. Eur. Am. M.	1,150,195	73,381	6.38	9,163	0.80	81,619	7.10	1,068,576	4.46
Non-His. Af.. Am. F.	280,263	24,505	8.74	4,628	1.65	28,579	10.20	251,684	7.12
Non-His. Af. Am. M.	289,326	25,196	8.71	4,892	1.69	29,474	10.19	259,852	9.38
Mex. Am. F.	296,109	24,681	8.34	5,956	2.01	29,873	10.09	266,236	3.47
Mex. Am. M.	307,424	25,736	8.37	6,215	2.02	31,160	10.14	276,264	4.26

Birth Cohort	high school and below (z=0)				college and above (z=1)			
	# of Births	# of Deaths	CDR	mean (g)	# of Births	# of Deaths	CDR	mean(g)
Non-His. Eur. Am. F.	432034	2044	4.73	3278	583889	1429	2.45	3390
Non-His. Eur. Am. M.	454249	2889	6.36	3393	614327	1876	3.05	3512
Non-His. Af.. Am. F.	162394	1254	7.72	3061	89290	539	6.04	3150
Non-His. Af. Am. M.	167571	1712	10.22	3165	92281	725	7.86	3266
Mex. Am. F.	222480	795	3.57	3299	43746	130	2.97	3322
Mex. Am. M.	230989	1028	4.45	3393	45275	150	3.31	3424

Non-His. = non- Hispanic Eur. = European Af. = African Mex. = Mexican

Am. = American F. = females M. = males

bwt = birth weight (gram)

*: missing birth weight, birth weight < 500 grams, missing LMP gestational age, LMP gestational age < 20 weeks

&: missing education information

CDR = Crude death rate (deaths per 1000 births)

Table 2 Definitions of the CDDmlr model with an indicator variable as the covariate

Symbol	Definition
Mixture Submodel parameters for the i subpopulation (i = s and p)	
--- functions of dichotomous covariate z	
$\pi_s(z)$	Mixing proportion (% secondary subpopulation)
α_0	Constant for z=0
α_1	Additive effect on the constant when z=1
$\mu_i(z)$	Mean birth weight
$\gamma_{i,0}$	Constant for z=0
$\gamma_{i,1}$	Additive effect on the constant when z=1
$\sigma_i(z)$	Standard deviation of birth weight
$\lambda_{i,0}$	Constant for z=0
$\lambda_{i,1}$	Additive effect on the constant when z=1
Mortality submodel parameters for the i subpopulation (i = s and p)	
--- coefficients of a second degree polynomial	
a^*_{i0}	Constant for z=0
b^*_{i0}	Linear term for standardized birth weight (x^*_i) for z=0
c^*_{i0}	Square term for standardized birth weight (x^*_i) for z=0
a^*_{i1}	Additive effect on the constant when z=1
b^*_{i1}	Additive effect on the linear term when z=1
c^*_{i1}	Additive effect on the square term when z=1

Table 3 Changes in parameter estimates for the mixture submodel due to higher education

Birth Cohort	$logit(\pi_s)$	π_s (%)	μ_s (g)	σ_s (g)	μ_p (g)	σ_p (g)
European American Females	-0.07	0.4	253	-51	103	-20
European American Males	-0.12 *	0.8 *	227	-66	108	-19
African American Females	-0.09	0.8	-65 *	114	88	-9
African American Males	-0.19	1.6	-136	152	95	9
Mexican American Females	-0.05 *	0.3 *	-36 *	-24 *	27	-3
Mexican American Males	0.03 *	-0.2 *	4 *	-29 *	35	-4

* not significant based on bias-adjusted 95% confidence intervals

Table 4 Decomposition of direct and indirect effects (expressed as log odds ratio) due to higher education

Birth Cohort	Primary			Secondary			Relative Contributions	
	Total	Indirect	Direct	Total	Indirect	Direct	Primary	Secondary
European American Females	-0.85 *	0.06	-0.91 *	-0.19	-0.42	0.23	-0.80	-0.01
European American Males	-0.87 *	-0.06	-0.81	-1.30 *	0.16	-1.46 *	-0.80	-0.09
African American Females	-0.66 *	0.12 *	-0.78 *	-0.43	0.20	-0.63	-0.59	-0.04
African American Males	-0.52 *	0.02	-0.54 *	-0.47	1.48 *	-1.95 *	-0.47	-0.05
Mexican American Females	-0.14	0.45 *	-0.59 *	-0.98 *	0.53	-1.51 *	-0.13	-0.07
Mexican American Males	-0.21 *	-0.17 *	-0.04	-0.14	0.49 *	-0.62	-0.19	-0.01

*: significantly different from 0.0 based on bias-adjusted 95% confidence intervals

"Relative Contributions" are the total primary and secondary effects multiplied by the proportion of primary and secondary births, that is the relative size of the primary effect compared to the secondary effect with respect to the total change in mortality due to education.