Age at Onset of Obesity among U.S. Adolescents: Findings from the National Longitudinal Study of Adolescent Health *

Aiko Hattori, Ph.D.

University of California, Los Angeles

Chirayath M. Suchindran, Ph.D.

University of North Carolina at Chapel Hill

* Please note that this is a preliminary report of the study. Further work is planned (as discussed in the discussion section) and the results will be included in the final version of this paper to be uploaded on the PAA 2012 program website by April 2, 2012.

SUMMARY

Adolescence obesity is a growing public health concern in the United States and research on age at onset of obesity is required to inform policies as to the timing and target population of interventions. We use the National Longitudinal Study of Adolescent Health (Add Health) data to examine the differentials in age at onset. Use of Add Health poses many methodological challenges. Assessment of obesity was done only at four discrete time points; data gathering involves informative sampling and clustered sampling. In addition we postulate that an unknown proportion of the population may be long-term survivors of obesity or not at risk of occurrence of obesity. We use non-parametric maximum likelihood procedures to estimate the proportion obese by age in overcoming the data difficulties. In addition we use accelerated failure time models to estimate the distribution of time to obesity in the presence of long-term survivors of obesity. The results suggest that non-Whites and males are more likely to become obese earlier compared with non-Hispanic Whites and females, respectively, when at risk for obesity. However, non-Hispanic Whites and males have lower risk for obesity, compared with non-Whites and females, respectively. Health policies should promote preventive efforts as early as adolescence and target high risk groups, including non-white and both girls and boys.

INTRODUCTION

Obesity among adolescents is a pressing public health problem in the United States. It is estimated that the proportion of obese adolescents (i.e., BMI equal to or greater than the 95th percentile of the sex-specific BMI growth charts) ages 12-19 increased from 5.0% to 18.1% between 1976-1980 and 2007-2008 (Ogden and Carroll 2010). Obesity has negative consequences for physical, psychological, and social well-being throughout the life course, including increased risk for adulthood obesity (Dietz 1998; Goodman et al. 2000; Harris et al. 2009; Reilly et al. 2003; Serdula et al. 1993; Whitaker et al. 1997), comorbidities (Dietz 1998; Harris et al. 2009), and deteriorated social life (Goodman et al. 2000; Gortmaker et al. 1993; Harris et al. 2009; Strauss and Pollack 2003). Adolescence obesity is associated with disparities among socio-demographic subgroups including gender and race/ethnicity (Goodman et al. 2000; Harris et al. 2009; Ogden and Carroll 2010; The et al. 2010), in which males and racial/ethnic minorities are more likely to be overweight or obese than their respective counterparts. Research on time to obesity among adolescents is required to inform policies as to the timing and target population of interventions.

Survival analysis, a statistical method to examine time to event of interest and its association with risk factors, have been widely adopted to examine time to obesity. However, it faces methodological challenges when analyzing complex survey data, including the National Longitudinal Study of Adolescent Health (Add Health). Major methodological challenges lie in that: 1) data collected through multistage sampling survey are often clustered and have unequal sampling probabilities associated with time to obesity; 2) the exact time to obesity may not be observed by researchers; and 3) there may be an unknown subpopulation not at risk for obesity.

Multistage sampling designs often cause observations at a lower stage of sampling to be correlated with each other at a higher stage of sampling. When the clustering is ignored in data analyses, variance estimates of parameters are often biased downward. Multistage sampling often cause sampling probabilities to vary across a sample. When sampling probabilities at one or more sampling stages depend on time to obesity, the sampling design is said to be informative. Informative sampling design causes standard estimators of parameters to be biased (Grilli and Pratesi 2004; Korn and Graubard 2003; Pfeffermann et al. 1998; Skinner 1989). The direction and magnitude of the bias cannot be determined as a priori knowledge.

An example can be drawn from Add Health. Add Health is a school-based multistage longitudinal survey of adolescents started in 1994, applying a probability proportionate to size design. In assessing time to obesity, the school sizes may be associated with the obesity status. For instance, school sizes may be associated with human and financial capacities supposedly associated with health promotion programs, such as healthy meal plans and physical exercise, implying that the sampling design is informative.

Time to event data collected through panel surveys, including Add Health, often involve interval censored data. When the exact time to obesity is not observed by a researcher, the event times are said to be censored. Specifically, event times are right-censored if study participants are lost to follow-up or do not become obese by the end of the study, left-censored if they are already obese by the first observation time point, and interval-censored if they become obese between two observation points and the exact time is not known. While right-censored event times can be accounted for by both Accelerated Failure Time (AFT) and Proportional Hazard (PH) models, interval-censored event times require model specification within the framework of AFT model (Allison 1995).

The standard survival analysis assumes that every individual is at risk for obesity, of which assumption may not hold when there is a subpopulation not at risk. Under a fixed period time for observation or data collection, right-censored observations are common in survey data. Right-censored observations may be comprised of two different subpopulations: at-risk individuals whose survival time exceeded the last observation time point, and “long-term survivors” not at risk for obesity. The subpopulation not at risk for obesity is often not known to the researcher and cannot be distinguished from at-risk individuals whose survival time exceeded the study period. Consequences of the inappropriate assumption include biased estimates of parameters (Taylor et al. 2003). In this paper we present an approach to addressing the two groups of right-censoring based on the assumption that there is an unknown proportion of subjects who can be classified as long-term survivors of obesity. The assumption is plausible because the majority of adolescents survive their adolescence without becoming obese (Ogden and Carroll 2010).

In summary, the three methodological challenges, i.e., clustered data with informative sampling probabilities, interval-censored event time, and an unknown subpopulation not at risk for obesity, suggest that parameter estimates may be biased and hypothesis testing results may be invalid if statistical analysis fails to account for them. In contrast to the increasing application of survival analysis to research on time to obesity, there is little research addressing the methodological problems. This study demonstrates a statistical approach to addressing the three challenges by extending the PML approach to a random effect model with a mixture distribution within the framework of AFT model. We examine time to obesity among adolescents and young adults and its association with major demographic characteristics including gender and race/ethnicity by using national representative longitudinal data from Add Health.

METHODS

Data

Add Health is a nationally representative, school-based longitudinal study of adolescents aimed at exploring health related behaviors (Gordon-Larsen et al. 2004) and conducted by the Carolina Population Center, University of North Carolina at Chapel Hill. Data were collected from students in grades 7 to 12 enrolled in schools thorough in-home interview. Study details are available elsewhere(Carolina Population Center. The University of North Carolina at Chapel Hill). At the first wave in 1994 -1995, 20,745 adolescents in grades 7 through 12 (ages 12-20 years) were interviewed and followed up in 1996 (ages 13-21 years, N=14,738), in 2001-2002 (ages 18- 26 years, N=15,197), and in 2007-2009 (ages 24-32 years, N=15,701). At wave I, schools and adolescents were sampled based on a multistage, proportionate to size sampling involving oversampling of some racial/ethnic groups. The analytical sample used in this study excludes respondents with missing data and is comprised of 18,466 respondents from 130 schools.

Descriptive statistics of the respondents, weighed by their sampling probability, are presented in Table 1. Males and females are distributed almost evenly (51.0% and 49.0%, respectively). The majority of respondents self-classified their race/ethnicity as non-Hispanic Whites (72.7%), followed by African-Americans (16.0%). Of the 18,466 respondents, 2,059 (11%) were left-censored, 12,549 (68%) were right-censored, 3,848 (21%) were interval-censored.

Table 1. Descriptive statistics of respondents (N=18,466)

	%
Gender	
Male	51
Female	49
Race/Ethnicity	
Non-Hispanic White	72.6
Other	27.4

Note: observations are weighted based on the final sampling probability

Measurements

Body mass index (BMI, calculated as weight in kilograms divided by square of height in meters) is used as a measure of relative weight status. BMI values of 30 or greater or age- and sex-adjusted BMI percentiles of 95 or greater based on the growth charts recommended by the Center for Disease Control and Prevention (Ogden et al. 2002; The et al. 2010) are classified as obese. There are 71 respondents at wave III whose weight exceeded the scale capacity. These respondents are classified as obese. Time to obesity is measured in respondent's age in years.

Statistical Analyses

We first estimate the survival curves using a non-parametric method proposed by Turnbull (Refernce of Turnbull paper) by incorporating sampling weights. We further extend the PML method to a random effect, mixture distribution model within the framework of AFT model to assess time to obesity among adolescents using the Add Health data. The model addresses the survey's potentially informative sampling design, clustered data, censored event times, and an unknown subpopulation not at risk for obesity. Analytical details are provided in the appendix. Tests were performed through the regression model and all results were considered statistically significant at P<0.05 unless indicated otherwise. For all analyses, SAS version 9.2 (SAS Institute, Cary, NC) is used.

RESULTS

Figure 1 plots the survival curve (i.e., the cumulative proportion not obese by age) for all the respondents using the non-parametric procedure proposed by Turnbull (1976). The estimated cumulative probability of obesity by age 33 is 0.47.

Figures 2 and 3 plot survival curves stratified by race/ethnicity (i.e., non-Hispanic Whites vs. others) and gender (males vs. females), respectively. Non-Hispanic Whites have a higher survival probability until their early 30's, after which the gap in the survival probabilities closes between non-Hispanic Whites and non-Whites (Figure 2). The estimated probability of obesity by age 33 is 0.51 for non-Hispanic Whites and 0.49 for non-Whites, respectively. Males and females exhibit similar survival curves during adolescences and early adulthoods. The estimated probability of obesity by age 33 is 0.47 for both males and females.

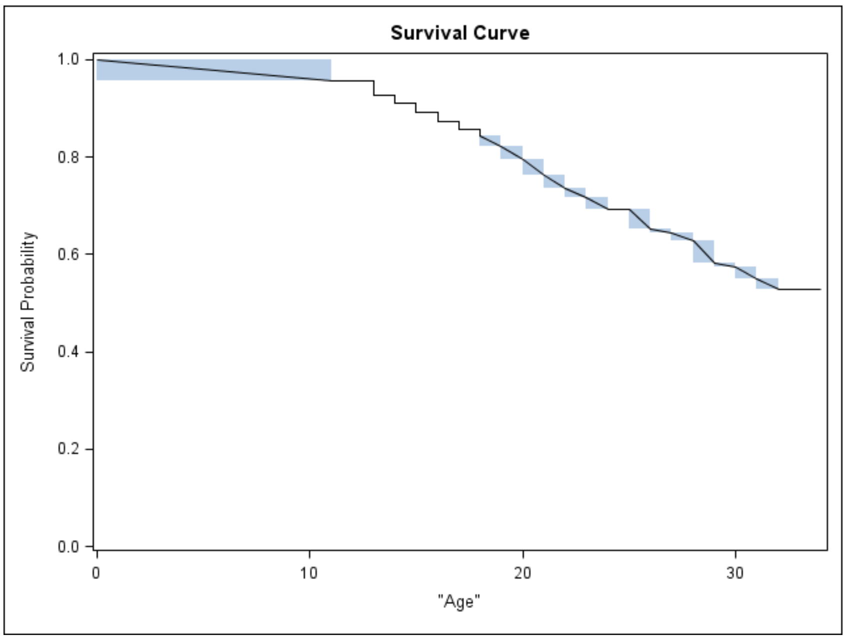

Figure 1. Survival curve of all the respondents

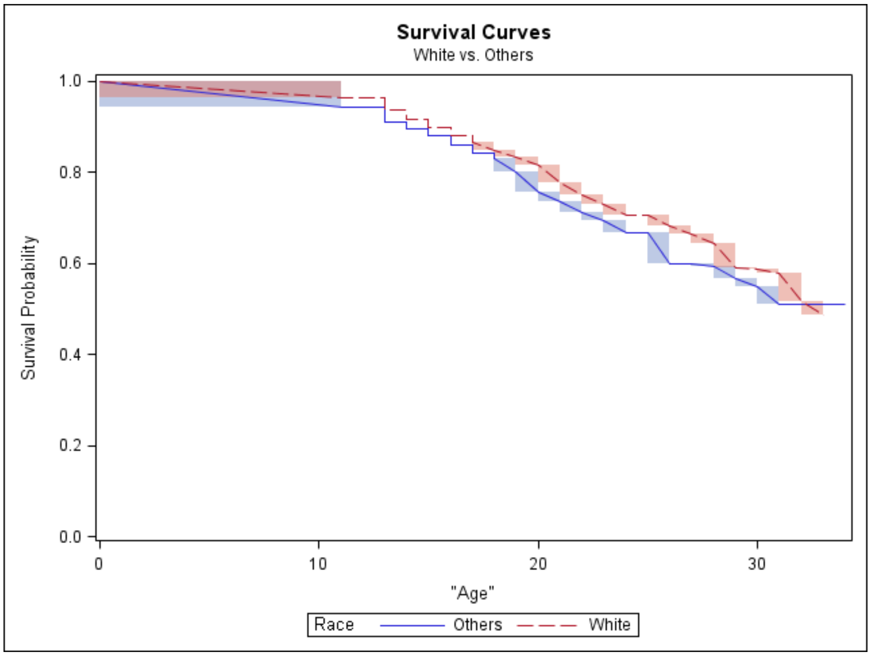

Figure 2. Survival curve by race/ethnicity

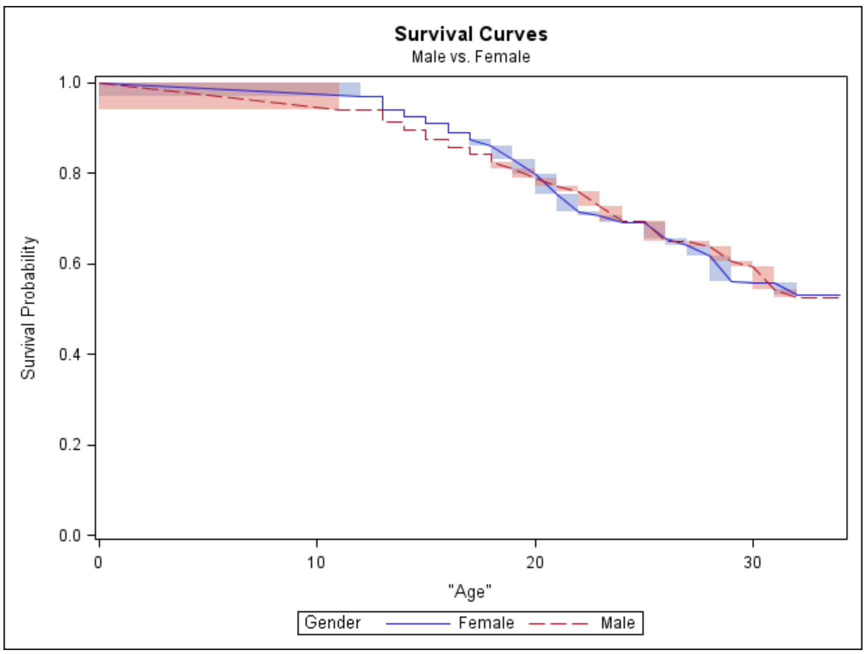

Figure 3. Survival curve by gender

Table 2. Survival analysis results

	Unadjusted model (LIFEREG procedure)			Mixture distribution model (NLMIXED procedure)		
	Coefficient estimate	(SE)	Pr(Z>|z|)	Coefficient estimate	(SE)	Pr(Z>|z|)
Time to Obesity						
Intercept	3.533	0.0189	<.001	3.278	0.0098	<.001
White	0.148	0.0196	<.001	0.063	0.0121	<.001
Males	0.109	0.0162	<.001	-0.130	0.0087	<.001
Interaction	-0.116	0.0192	<.001	-0.032	0.0162	0.048
Probability of Obesity						
Intercept	-			0.971	0.0195	<.001
White	-			-0.390	0.0231	<.001
Males	-			-1.013	0.0269	<.001
Interaction	-			0.396	0.0271	<.001
Scale	0.447	0.0091		0.356	0.0100	
Theta	-			0.016		

Results from the Unadjusted Model

The estimated average time to obesity obtained through the unadjusted model is 34.2 years (i.e., $\exp(3.533)$) among non-White females. It is 16.0% (i.e., $100 \times \{\exp(0.148) - 1\}$) longer among non-Hispanic White females and 11.5% (i.e., $100 \times \{\exp(0.109) - 1\}$) longer among non-White males than among non-White females. Overall, non-Hispanic White males have an average time to obesity 15.1% longer than non-White females. The corresponding hazard of obesity is 72 % (of the hazard for non-White females) for non-Hispanic White females; and 78 % (of the hazard for non-White females) for non-White males is 78 %, and 73 % (of the hazard for non-White females) for non-Hispanic White males.

Results from the Mixture Distribution Model

The results of the mixture distribution model are distinct from the model unadjusted for the aforementioned statistical assumptions. The estimated time to obesity among non-White females at risk for obesity is 26.5 years (i.e., $\exp(3.278)$), which is shorter than the corresponding estimate of 34.2 years obtained from the unadjusted model. The estimated difference of 7% in time to obesity between non-Hispanic Whites and non-Whites at risk for obesity is also smaller than the corresponding estimate of 16%. The study conclusion as to the association between gender and time to obesity is reversed, given the negative coefficient estimate of -0.130; the average time to obesity is 12.2% shorter among non-White males compared with non-White females at risk for obesity. Overall, non-Hispanic White males at risk have an average time to obesity 9.4% shorter than that among other race/ethnicity females at risk.

At the same time, the results suggest that there may be a subpopulation not at risk for obesity and that non-Hispanic Whites and males have lower risk for obesity, compared with non-Whites and females, respectively. The estimated probability of obesity among non-Hispanic White females is 0.641 (i.e., **Error! Bookmark not defined.** $1 - 1/\{1 + \exp(0.971 - 0.390)\}$) while that among non-White females is 0.725 (i.e., $1 - 1/\{1 + \exp(0.971)\}$). Likewise, the estimated probability of obesity among non-White males is 0.490 (i.e., $1 - 1/\{1 + \exp(0.971 - 1.013)\}$). Overall, non-Hispanic White males have a probability of 0.491 (i.e., $1 - 1/\{1 + \exp(0.971 - 0.391 - 1.013 + 0.396)\}$).

DISCUSSION

Survival analysis is widely used to assess time to obesity using complex survey data. It may face methodological challenges when analyzing complex survey data: clustered data with informative sampling probabilities, interval-censored observations, and an unknown subpopulation not at risk. This study addresses the three challenges by extending the PML method to a random effect, mixture distribution model within the framework of AFT model. The study results confirm that a model failing to address the problems may produce biased estimates of parameters.

The unadjusted model assuming risk for obesity for all the population identify non-Whites and females as at increased risk of becoming obese earlier compared with their respective counterparts. On the other hand, the mixture distribution model provides different conclusions especially as to the association of covariates with time to obesity and probability of obesity. Non-Hispanic White adolescents are found to have a longer time to obesity and lower risk for obesity compared with their non-White counterparts. Males have a shorter time to obesity when at risk for obesity but have lower risk for obesity compared with females. They may imply that there are more females than males at risk for obesity, but males at risk for obesity are likely to become obese earlier than females at risk. When the two different subpopulations – one at risk and the other not at risk for obesity – are not addressed in a model, the regression coefficients may be biased and obscure true association between covariates and time to obesity.

The study results have policy implications in addressing obesity and its health consequences. Given that approximately 32% of respondents were found obese by the end of the study, public health policy should promote preventive and curative efforts as early as adolescence and young adulthood. Also public health policy should target high risk groups, including non-white and both girls and boys, to address the disproportionate prevalence and risk of obesity. At the same time, the mechanisms through which these demographic characteristics are associated with obesity should be investigated further. For instance, these demographic characteristics may represent other factors, including biological, cultural, environmental, behavioral, and socio-economic factors. To develop effective interventions to address high risk groups, future research should look into the mechanisms by collecting information on these factors and assessing their relations with obesity.

This study has several limitations. First, the anthropometric information collected at wave I is based on self-reporting, unlike that at waves II, III, and IV, which is measured by trained interviewers. Therefore it is possible that the anthropometric information at wave I contains reporting errors. To examine reporting errors of self-reported weight and weight among adolescents and their impact on results of studies assessing adolescent obesity, Goodman *et al*. (2000) compare obesity status between waves I and II, using the Add Health data. They conclude that 96% of respondents are correctly classified as to obesity status when self-reported weight and height are used to calculate BMI. Because they assess the same data used in this study, their conclusion may be generalized to this study, suggesting that the impact of reporting errors on the study results may be minimal if any at all.

Second, the probability estimation is available only for Turnbull intervals. While the estimation method is well suited for the Add Health data because the respondents were observed in overlapping intervals, the method may not be applicable to other data if the intervals are fixed or non-overlapping. Therefore, exploration of data and careful application of the method is required.

Future Work

We plan to refine the mixture distribution model and the results will be included in the final draft of this paper to be uploaded on the PAA 2012 website by April 2, 2012. In this preliminary report, we assumed that both gender and race/ethnicity are associated with both time to obesity and risk for obesity. However, it is possible that different sets of covariates may be associated with the two outcomes. Future work therefore includes exploration and selection of covariates into the different components of distributions, i.e., Weibull and Bernoulli distributions, through AIC and BIC.

ACKNOWLEDGEMENTS

This research uses data from Add Health, a program project directed by Kathleen Mullan Harris and designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill, and funded by grant P01-HD31921 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 23 other federal agencies and foundations. Special acknowledgment is due Ronald R. Rindfuss and Barbara Entwisle for assistance in the original design. Information on how to obtain the Add Health data files is available on the Add Health website (http://www.cpc.unc.edu/addhealth). No direct support was received from grant P01-HD31921 for this analysis.

REFERENCES

Allison, P.D. 1995. *Survival analysis using SAS : a practical guide*. Cary, NC: SAS Institute.

Carolina Population Center. The University of North Carolina at Chapel Hill. "Add Health." 2011(09/14).

Dietz, W.H. 1998. "Health Consequences of Obesity in Youth: Childhood Predictors of Adult Disease." *Pediatrics* 101:518-525.

Goodman, E., B.R. Hinden, and S. Khandelwal. 2000. "Accuracy of Teen and Parental Reports of Obesity and Body Mass Index " *Pediatrics (Evanston)* 106(1):52-58.

Gordon-Larsen, P., L.S. Adair, M.C. Nelson, and B.M. Popkin. 2004. "Five-year obesity incidence in the transition period between adolescence and adulthood: the National Longitudinal Study of Adolescent Health " *The American Journal of Clinical Nutrition* 80(3):569.

Gortmaker, S.L., A. Must, J.M. Perrin, A.M. Sobol, and W.H. Dietz. 1993. "Social and Economic Consequences of Overweight in Adolescence and Young Adulthood " *The New England Journal of Medicine* 329(14):1008-1012.

Grilli, L. and M. Pratesi. 2004. "Weighted estimation in multilevel ordinal and binary models in the presence of informative sampling designs." *Survey Methodology* 30(1):93-103.

Harris, K.M., K.M. Perreira, and D. Lee. 2009. "Obesity in the Transition to Adulthood: Predictions Across Race/Ethnicity, Immigrant Generation, and Sex " *Archives of Pediatrics & Adolescent Medicine* 163(11):1022-1028.

Korn, E.L. and B.I. Graubard. 2003. "Estimating Variance Components by Using Survey Data." *Journal of the Royal Statistical Society.Series B (Statistical Methodology)* 65(1):pp. 175-190.

Ogden, C.L. and M.D. Carroll. 2010. "Prevalence of Overweight, Obesity, and Extreme Obesity Among Adults: United States, Trends 1976–1980 Through 2007–2008 " .

Ogden, C.L., R.J. Kuczmarski, K.M. Flegal, Z. Mei, S. Guo, R. Wei, L.M. Grummer-Strawn, L.R. Curtin, A.F. Roche, and C.L. Johnson. 2002. "Centers for Disease Control and Prevention 2000 Growth Charts for the United States: Improvements to the 1977 National Center for Health Statistics Version." *Pediatrics (Evanston)* 109(1):45-60.

Pfeffermann, D., C.J. Skinner, D.J. Holmes, H. Goldstein, and J. Rasbash. 1998. "Weighting for unequal selection probabilities in multilevel models " *Journal of the Royal Statistical Society.Series B, Statistical Methodology* 60(1):23-40.

Reilly, J.J., E. Methven, Z.C. McDowell, B. Hacking, D. Alexander, L. Stewart, and C.J.H. Kelnar. 2003. "Health consequences of obesity " *Archives of Disease in Childhood* 88(9):748-752.

Serdula, M.K., D. Ivery, R.J. Coates, D.S. Freedman, D.F. Williamson, and T. Byers. 1993. "Do obese children become obese adults? A review of the literature " *Preventive Medicine* 22(2):167.

Skinner, C.J. 1989. "Domain means, regression and multivariate analysis " in *Analysis of Complex Surveys* , edited by C.J. Skinner, D. Holt, and T.M.F. Smith. Chichester: Wiley.

Strauss, R.S. and H.A. Pollack. 2003. "Social Marginalization of Overweight Children " *Archives of Pediatrics & Adolescent Medicine* 157(8):746-752.

Taylor, D.J., M.A. Weaver, and R.E. Roddy. 2003. "Evaluating factors associated with STD infection in a study with interval-censored event times and an unknown proportion of participants not at risk for disease " *Statistics in Medicine* 22(13):2191-2204.

The, N.C., C. Suchindran, K.E. North, B.M. Popkin, and P. Gordon-Larsen. 2010. "Association of Adolescent Obesity With Risk of Severe Obesity in Adulthood " *JAMA : The Journal of the American Medical Association* 304(18):2042-2047.

Turnbull, B.W. 1976. "The Empirical Distribution Function with Arbitrarily Grouped, Censored and Truncated Data." *Journal of the Royal Statistical Society.Series B (Methodological)* 38(3):pp. 290-295.

Whitaker, R.C., J.A. Wright, M.S. Pepe, K.D. Seidel, and W.H. Dietz. 1997. "Predicting Obesity in Young Adulthood from Childhood and Parental Obesity " *The New England Journal of Medicine* 337(13):869-873.

APPENDIX

Non-Parametric Estimation of Survival Curves

Let T_k denote the time to obesity for the k th subject for k =1, 2,… n. Let $(O_k, U_k]$ be the interval for which T_k is measured where O_{ij} and U_{ij} are lower and upper time points, respectively. Note that the interval $(O_k, U_k]$ may overlap. From the data of overlapping intervals $(O_k, U_k]$, Turnbull (1976) derived a procedure to generate a set of non-overlapping intervals $\{q_l, p_l\} = \{(q_1, p_1],...,(q_m, p_m]\}$ for $l = 1,...,m$, over which the survival curve S(t) = $\mathbf{Pr}(T > t)$ can be estimated (Turnbull 1976). The set of non-overlapping intervals are obtained from all left, interval and right censored intervals in such a way that for q_l is a left end point and p_l is the right end point and there is no other left or right end point between q_l and p_l.

Turnbull (1976) proposed an EM algorithm to obtain the probability distribution (as well as the survival function). The procedure can be briefly summarized in the following way. Let $\theta_l = \Pr\left(q_l < T < p_l, l = 1,...,m\right)$. Note that $\sum_{l=1}^{m} \theta_l = 1$. The iterative estimation of the parameters θ_l starts with initial estimates, usually assumed to be uniform in the intervals.

The revised estimates are calculated as:

$$\hat{\theta}_k(t_l) = \frac{\hat{\theta}(t_l) I\{t_l \in (O_k, U_k]\}}{\sum_{t_l \in (O_k, U_k)} \hat{\theta}(t_k)}.$$

Then in the maximization step, an improved estimate can be obtained as $\hat{\theta}(t_l) = \frac{1}{n}\sum_{k=1}^{n} \hat{\theta}_k(t_l)$ and the process is continued until convergence. In order to account for sampling weights, a weighted average is calculated in the maximization step.

Random Effect, Mixture Distribution Model

To address the aforementioned methodological challenges, i.e., clustering, informative sampling weights, and an unknown subpopulation not at risk, we introduce a frailty term, two levels of sampling weights, and a mixture distribution, respectively. They are modeled within the framework of AFT model to account for interval-censored event times.

Suppose observations from school i are independent conditional on an unobserved frailty (b_i). For computational simplicity we assume the frailties follows a normal (0, θ) distribution. We assume that time to obesity follows a Weibull (λ, p) distribution. Let $\mathbf{x}_{ija}$ be a vector of covariates (e.g., race/ethnicity and gender) for j th respondent from i th school, $\boldsymbol{\alpha}$ be a vector of corresponding coefficients, and $\lambda = \exp(-\alpha_0)$.

Suppose the sampling is done as follows: At the first stage N schools are selected with the inclusion probability π_i, $i = 1,..., N$. At the second stage n_i respondents are selected within i th selected school with the inclusion probability $\pi_{j|i} (i = 1,..., n_i)$. Define the first stage sampling weight $w_i = 1/\pi_i$ and the second stage sampling weight $w_{j|i} = 1/\pi_{j|i}$, respectively. Then $w^*_{j|i} = (w_{j|i} \times n_i)/\sum_j w_{j|i}$.

Finally, assume that right-censored observations consist of two groups: adolescents at risk for obesity but did not become obese by the last observation point, and those not at risk for obesity. Let Y be a random variable representing risk for obesity. That is, Y_{ij} =1 if respondent j from school i is at risk and Y_{ij} =0 otherwise. Then the probability of being at risk can be written as (Taylor et al. 2003):

$$\Pr\{Y_{ij} = 1 \mid \mathbf{x}_{ij2}\} = \exp(\mathbf{x}_{ij2}\boldsymbol{\beta}) / \{1 + \exp(\mathbf{x}_{ij2}\boldsymbol{\beta})\},$$

where $\mathbf{x}_{ij2}$ is a vector of covariates, which may be different from $\mathbf{x}_{ij1}$ (the vector of covariates associated with time to obesity) and $\boldsymbol{\beta}$ is a vector of corresponding coefficients. Suppose there are n_{i1} respondents who were not right-censored and n_{i2} respondents who were right-censored from school i for $n_i = n_{i1} + n_{i2}$.

Then the estimate of census likelihood can be specified as follows:

$$\log \hat{L}(\boldsymbol{\alpha},\theta,p)=\sum_{i=1}^{N} w_i \log \int \left\{ \exp\left[\sum_{j=1}^{ni1} w_{j|i}^{*} \log\left(\Pr(Y_{ij}=1 \mid \mathbf{x}_{ij2}) L_{ij}(\boldsymbol{\alpha},\theta,p \mid b_i) \right) \right] \right\} \phi(b_i)\,db_i$$

$$+\sum_{i=1}^{N} w_i \log \int \left\{ \exp\left[\sum_{j=1}^{ni2} w_{j|i}^{*} \left(\Pr(Y_{ij}=0 \mid \mathbf{x}_{ij2}) + \log\left(\Pr(Y_{ij}=1 \mid \mathbf{x}_{ij2}) L_{ij}(\boldsymbol{\alpha},\theta,p \mid b_i) \right) \right) \right] \right\} \phi(b_i)\,db_i$$

where

$$L_{ij}(\boldsymbol{\alpha},\theta,p)=\begin{cases} \exp\left\{-\left[O_{ij}\exp(-\mathbf{x}_{ij1}\boldsymbol{\alpha})+b_i\right]^{1/p}\right\},\ if\ U_{ij}=\infty, \\ 1-\exp\left\{-\left[U_{ij}\exp(-\mathbf{x}_{ij1}\boldsymbol{\alpha})+b_i\right]^{1/p}\right\},\ if\ O_{ij}=0, \\ \exp\left\{-\left[O_{ij}\exp(-\mathbf{x}_{ij1}\boldsymbol{\alpha})+b_i\right]^{1/p}\right\}-\exp\left\{-\left[U_{ij}\exp(-\mathbf{x}_{ij1}\boldsymbol{\alpha})+b_i\right]^{1/p}\right\},if\ O_{ij}\neq U_{ij}, \\ \lambda p(\lambda O_{ij})^{(p-1)}\exp(-\mathbf{x}_{ij1}\boldsymbol{\alpha}+b_i)^{1/p},if\ O_{ij}=U_{ij}. \end{cases}$$

The estimated census likelihood function accounts for clustering, informative sampling weights, interval-censored event times, and an unknown subpopulation not at risk for obesity. The NLMIXED procedure of SAS is well suited for the analysis purposes because it can easily incorporate these model specifications. We compare the parameter estimates with those obtained from a model unadjusted for the clustering, informative sampling weights, and an unknown subpopulation not at risk of obesity. Differences in the parameter estimates, therefore, can be interpreted as the net impact of the addressed statistical assumptions. Variances are estimated through a jackknife procedure.