ABSTRACT

Title of Dissertation: LATENT CLASS LOGISTIC REGRESSION
WITH COMPLEX SAMPLE SURVEY DATA

Steven A. Blahut, Doctor of Philosophy, 2004

Dissertation directed by: Professor C. Mitchell Dayton
Department of Measurement, Statistics, &
Evaluation

Latent class regression has been reported previously in the literature. Often, however, data are collected from a survey that utilizes unequal selection probabilities that result in complex sample survey data. Techniques for latent class logistic regression utilizing complex survey data have not previously been reported. Additionally, no software is available to perform these analyses. A model was chosen for investigation based on an existing survey called the Indiana Youth Tobacco Survey. A variety of scenarios were investigated using systematically manipulated conditions to simulate complex sample survey data. Specifically, the effect of ignoring sample weights was investigated by comparing bias in parameter estimates from simulations both incorporating and ignoring weights. Additionally, several competing approaches for estimating standard errors were compared in terms of bias and confidence interval coverage. The techniques that were investigated were the unadjusted approach

assuming simple random sampling, the jackknife, the bootstrap, and the design effect adjustment. Two design effects were compared, one based on jackknife estimates and one based on bootstrap estimates. The results indicated that weights must be incorporated in the estimation via pseudo-maximum likelihood to ensure that parameter estimates are not biased. These estimates were less biased than jackknife, bootstrap, and unweighted parameter estimates. In terms of variance estimation, the bootstrap estimates were preferred. Estimates arising from the assumption of simple random sampling were consistently small and therefore undesirable. Jackknife and design effect adjusted standard errors were better, but bootstrap standard errors were consistently best. Finally, the best technique was applied to the Indiana Youth Tobacco Survey data to identify latent classes that differed in their susceptibility to initiate tobacco use and abuse. The results indicated that a two class model was a better fit to the data than a one class model. These classes differed in their susceptibility to peer pressure. Latent class one comprised 82% of the population and was more susceptible to peer pressure than was latent class two. Both classes were more at risk of initiating tobacco use as they aged.

LATENT CLASS LOGISTIC REGRESSION WITH COMPLEX SAMPLE SURVEY DATA

By

Steven A. Blahut

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2004

Advisory Committee:

Professor C. Mitchell Dayton, Chair
Dr. Francis Alt
Dr. Amy Hendrickson
Professor George Macready
Professor Robert Mislevy

DEDICATION

This work is dedicated to my son, Donovan, my parents, John and Linda and most of all, my wife, Tara for their inspiration dedication, and support.

ACKNOWLEDGMENTS

I am quite grateful to Professor C. Mitchell Dayton, a gifted researcher and teacher for his insightful guidance and support. Additionally, I wish to thank the faculty and support staff in the EDMS department for making the past five years incredibly rewarding.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SAS PROGRAMS

CHAPTER I
PURPOSE AND RATIONALE

Background

Logistic regression (LR) has been a common analytic tool in epidemiology and social science for many years. In fact, it has become the analysis of choice for relating a categorical response or criterion variable to a set of predictor variables often referred to as covariates. In other areas of study, for example, marketing, the technique has been enhanced to incorporate unobservable or latent populations. This type of model is called latent class regression and has been developed for various regression models (Wedel & Desarbo, 2002). In these scenarios, the population may be characterized by latent subgroups that are best represented by differing regression equations. If the population is indeed heterogeneous, the mixture approach is preferred as it will identify these unobserved subgroups and lend insight into their differences. Quite often, the data analyzed by these and other similar methods arise from sampling techniques that are more complex than simple random sampling (SRS). Complex samples are often stratified with some segment/strata of the population being sampled at a disproportionate rate. Additionally, complex surveys may involve clustering, for example, schools within school districts, and school districts within states, and so on. If these data are analyzed by a "model based" approach, that is, as if they arose from a SRS, parameter estimates will almost certainly be biased and resulting standard errors will be underestimated. As a result, hypothesis testing will be too liberal. Therefore, a "design based" approach where clustering and stratification are taken into account during variance estimation is desirable. Often, sampling weights reflecting the

probability of selection, post stratification, and non-response are used during parameter estimation to reduce potential bias.

Logistic regression models with complex survey data have been developed and are described elsewhere in detail by Korn & Graubard (1991), Roberts, Rao, & Kumar (1987), Skinner, Holt, & Smith (1989), and Thomas & Rao (1987) among others. The basic idea as developed by Roberts, Rao, and Kumar (1987) is to approximate the likelihood function by incorporating the observed sample and the sampling weights to create a pseudo-likelihood. At this time, methods that incorporate complex sampling designs into latent class regression have not been rigorously developed or empirically validated.

The motivation for the research can be traced to the Indiana Youth Tobacco Survey (IYTS). This survey is administered annually to a representative sample of middle and high school-aged children in public schools in the state of Indiana. The specific problem of interest was trying to identify unobservable subgroups of students in relation to their susceptibility to initiating tobacco use.

Purpose of Study

The goal of this study was to investigate the problems of obtaining valid parameter estimates and standard errors for latent class logistic regression (LCLR) when analyzing complex sample survey data. To that end, differing approaches to parameter and variance estimation were investigated and results were based on samples generated to represent a contrived population with known characteristics. Specifically, the study had the following aims:

a) To investigate the importance of incorporating the sampling design when performing LCLR. That is, comparing "model based" parameters that ignore the sampling weights to "design based" results that incorporate weights.

b) To compare the performance of different variance estimators, specifically, the jackknife, bootstrap, and design effect adjustment. This was done by comparing simulated variances to empirically derived "true" variances.

These goals were achieved by Monte Carlo methods that simulated a variety of scenarios: two sample schemes (proportional and disproportionate over-sampling from strata), two clustering schemes (low and moderate clustering effect), and three latent class mixing proportions (ranging from moderate and extremely disparate). In each scenario, samples were simulated from two strata that were characterized by different latent class proportions.

Significance/Novelty of Study

While LR methods for complex survey data are widely accepted, techniques of parameter and variance estimation for LCLR with complex survey data have not been previously studied. It was hypothesized that ignoring sampling weights would have an adverse effect on parameter estimates and ignoring clustering would result in poor variance estimates. The original contributions of this study are:

(1) Methods for incorporating sampling weights and clustering into LCLR were developed and tested.

(2) Estimates of standard errors for LCLR parameters stemming from complex survey data were obtained through competing methods.

(3) SAS code written for the study makes these analyses more accessible to researchers in various disciplines.

The best performing procedure(s) were then applied to a latent class logistic regression analysis of the 2001 Indiana Youth Tobacco Survey (IYTS).

CHAPTER II
REVIEW OF THE LITERATURE

Finite Mixture Models

Traditionally the evolution of finite mixture models can be traced back to Pearson (1894) and Newcomb (1886). In finite mixture models, it is asumed that sample observations arise from a number of unobservable, latent classes of unknown proportion. The analysis of these models involves the fitting of mixtures (i.e. weighted sums) of distributions. These latent classes are characterized by distributional parameters that differ among classes. A variety of mixture models have appeared in the literature including mixtures of normal (Hasseblad, 1966), Poisson (Hasselblad, 1969), binomial (John, 1970), and exponential (Everitt & Hand, 1981) distributions among others. A comprehensive treatment of these and other mixture models is offered by Titterington, Smith, and Makov (1985).

Mixtures of Regressions

There are many scenarios in social science, education, biomedical, and physical science research when analysis involves the estimation of a linear model. The relation of a set of predictor variables to a criterion is one of the most widely practiced of all statistical techniques (McCullagh & Nelder, 1989; Pedhauzer, 1997; Kleinbaum, Kupper, & Muller, 1988). In many situations, it is likely that the estimation of one set of regression coefficients across all observations will lead to spurious results (Wedel & DeSarbo, 2002). This is particularly true if the data arise from several latent subgroups with different population parameters. This has lead to the development of latent class regression models. Latent class regression models involve the estimation of differing

regression coefficients for each latent class. These models have been developed for various types of data including normal (Desarbo & Cron, 1988; Quandt & Ramsey, 1978), count (Wedel et. al., 1991; Ramaswamy, Anderson, & DeSarbo, 1994), and binary (Follman & Lambert, 1989; Wedel & DeSarbo, 1992). The latter is the focus of this investigation.

The binary logistic regression model is nonlinear. The outcome variable π_i is the probability of having one of the two possible outcomes of the dichotomous variable y based on a nonlinear function of the best linear composite of predictors:

$$\pi_i = \frac{e^q}{1+e^q} \tag{2.1}$$

π_i is the model estimated probability of the *i*th case being in one of the categories and q is the familiar linear regression equation:

$$q = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_k X_k \tag{2.2}$$

with constant (intercept) β_0, regression coefficients β_j and predictors X_j for predictors (j = 1,2,…,k). A transformation of π_i, known as the logit (or log odds) transformation enables the resulting model to have the desirable properties of the linear regression model. This transformation is defined, in terms of π_i, as:

$$\ln\left(\frac{\pi_i}{1-\pi_i}\right) = \beta_0 + \sum \beta_j X_{ij} \tag{2.3}$$

The estimation procedure for coefficients is maximum likelihood, the goal of which is to identify the best combination of predictors to maximize the likelihood of obtaining the observed outcome frequencies. Maximizing the likelihood is tantamount

to maximizing the log-likelihood. The log-likelihood (LNL) is calculated by summing the probabilities for the predicted and actual outcome for each case:

$$LNL = \sum_{i=1}^{n} [y_i \ln(\pi_i) + (1 - y_i) \ln(1 - \pi_i)] \tag{2.4}$$

Since it is nonlinear, maximizing this function involves iterative computational methods such as Newton-Raphson.

The mixture of logistic regressions is a weighted sum of two or more components. The proportion of the mixture is denoted as θ and the sum of these proportions is equal to one. That is, $\sum \theta = 1$. Ordinary logistic regression is a special case where θ = 1.0. Therefore, for a two-component mixture, the resulting expression for the transformation of π_i is:

$$\ln\left(\frac{\pi_i}{1 - \pi_i}\right) = \theta\left(\beta_{01} + \sum \beta_{j1} X_{ij1}\right) + (1 - \theta)\left(\beta_{02} + \sum \beta_{j2} X_{ij2}\right) \tag{2.5}$$

And the resulting log likelihood is the same as (2.4) unless we introduce a vector of weights to perform pseudo-likelihood estimation (Hosmer & Lemeshow, 2000; Wedel & DeSarbo, 1992):

$$LNL = \sum_{i=1}^{n} [(w_h y_i) \ln(\pi_i) + (w_h (1 - y_i)) \ln(1 - \pi_i)] \tag{2.6}$$

Thus, for traditional maximum likelihood estimation, the weight vectors (for weights associated with each stratum h) w_h are simply a vector of ones. Follman and Lambert (1989) presented a study in which the two components (logits) comprising the mixture had the same slopes but different intercept constants. The current investigation used varying slopes and intercepts for each latent class.

Complex Sample Survey Data (CSSD)

Finite Population Correction (FPC)

If a sample of size *n* is taken from a larger population of size *N*, the sampling fraction is *n/N* (Lohr, 1999 p. 33). The finite population correction (FPC) denoted as $1 - n/N$ is required for variance estimators because the population is not infinite in size, and sampling is done without replacement. This correction is made because, with smaller populations, the sampling fraction is greater and thus more information is available about the population. In this case, variances are smaller. Typically, the FPC is ignored if the sampling fraction is $n/N < 0.05$, thus the FPC is close to one. Additionally, the FPC is ignored if inferences are being made to a larger super-population (Patterson, 1998). Therefore, the FPC was not utilized in the current investigation.

Stratification

Often, in survey research, there is supplementary information available to aid in the design of the sample. For example, gender, race, region, or other non-overlapping subgroups can aid in the creation of strata. In a population consisting of *H* strata ($h = 1, 2, \ldots H$), n_h observations are sampled from each stratum. For this approach to sampling to work, we assume that the summation of total observations in each stratum is equal to the total population. That is, $N_1 + N_2 + \ldots + N_H = N$. In proportional allocation, the number of sampled units from each stratum is proportional to their natural occurrence in the population. Thus, the probability of selection is n_h / N_h and is the same for all strata (Lohr, 1999). If not, it is referred to as disproportionate allocation (Cochran, 1977). If done appropriately, stratified sampling will result in more precise parameter estimates

(i.e. smaller variances). For the greatest possible precision, observations within each stratum should have very similar values and the stratum means should differ as much as possible.

To illustrate a stratified sample, consider a SRS paradigm to estimate a population mean of a dichotomous variable y taking on the values [0, 1]. Thus, the mean is a proportion and will be denoted P. The sample proportion, $\hat{P} = \frac{1}{n}\sum_{i \in S} y_i$. A proportion from a stratified sample is of the form:

$$\hat{P}_{str} = \sum_{h=1}^{H} \frac{N_h}{N} p_h \qquad (2.7)$$

Thus, the estimated population proportion is a weighted average of the sample stratum proportions. It is imperative that the size or relative size of the strata is known. The variance is:

$$\hat{V}(\hat{P}_{str}) = \sum_{h=1}^{H} \left(1 - \frac{n_h}{N_h}\right)\left(\frac{N_h}{N}\right)^2 \frac{P_h(1-P_h)}{n_h - 1} \qquad (2.8)$$

(Cochran, 1977; Lohr, 1999). The fpc for stratum *h* is $\left(1 - \frac{n_h}{N_h}\right)$. As previously noted, the fpc was dropped for this investigation, the resulting expression can be evaluated as if this quantity were equal to one.

Clustering

In cluster sampling, observations in the population are selected for inclusion in the sample if and only if they belong to a primary sampling unit (PSU). Observations or elements within a PSU are sometimes referred to as secondary sampling units (SSU). Typically, cluster samples are utilized for the sake of economy. The cost of sampling

SSUs is cheaper than that of each element in a SRS. Clusters, like strata, are a grouping of elements in a population. However, the selection of PSUs is quite different then strata. In one-stage cluster sampling, each element of the population falls into exactly one cluster, a SRS of clusters is selected and then all SSU in each cluster are sampled. In two-stage cluster sampling, a SRS of clusters is selected and then a SRS of SSU is taken from each PSU (Lohr, 1999). The current investigation utilized one-stage sampling with clusters of equal sizes as the PSU.

Where the utilization of stratified sampling is known to increase precision, cluster sampling leads to a reduction in precision. Observations within a given cluster tend to be more homogenous, thus reducing the effective sample size. For greatest precision, SSUs within a cluster should be heterogeneous and cluster means should be equivalent (Lohr, 1999, p. 133). The intraclass correlation coefficient (ICC), ρ ,is a conventional measure of intra-cluster homogeneity. The ICC is described elsewhere in detail and interested readers are referred to Lohr (1999, p.139-143). Clustering leads to a loss of precision when the ICC is positive. The loss of precision is effected by the magnitude of the ICC and the size of the PSU. Assuming a constant cluster size, M , an estimator of the population proportion P is:

$$\hat{P}_c = \frac{\hat{t}}{NM}, \tag{2.9a}$$

with

$$V(\hat{P}_c) = \left(1 - \frac{n}{N}\right)\frac{s_t^2}{nM^2} \tag{2.9b}$$

where $\hat{t}$ is the estimator of the total in the *i*th PSU,

$$\hat{t} = \frac{N}{n}\sum_{i \in S} t_i \tag{2.9c}$$

with sample variance s_t^2,

$$s_t^2 = \frac{1}{n-1}\sum_{i \in S}\left(t_i - \frac{\hat{t}}{N}\right)^2 . \qquad (2.9d)$$

A comprehensive development of the unequal cluster size scenario is given by Lohr (1999).

Weighting

Most large, nationally representative surveys such as NHANES (National Health and Examination Survey) and NYTS (National Youth Tobacco Survey) include sampling weights in public distribution data sets. These weights must be integrated into analyses so that parameter estimates are unbiased. When observations have the same probability of being selected for inclusion in a sample (such as in a SRS), they are not essential to ensure valid parameters and test statistics (Korn & Graubard, 1991; Lohr, 1999). These weights, when applicable, can be easily incorporated into ratio, mean, and total estimation (Cochran, 1977; Lohr, 1999).

Weights assign greater or less importance to some sampled observations over others. Weights are necessary when there is an unequal probability of selection. For example, in a stratified sample the sampling weight can be expressed as the inverse of the probability of selection:

$$w_{str} = \frac{N_h}{n_h} \qquad (2.10)$$

Thus, the sum of the sampling weights is equal to the total sample size, and each sampled observation is representative of a certain number of observations in the

population (Lohr, 1999). Weights in cluster samples are calculated in a different manner.

$$w_{clus} = \frac{NM}{nm} \qquad (2.11)$$

The use of these sampling weights in would result in the analysis of *N* observations. Therefore, relative weights are often used in the analysis of survey data, so results are based on *n* observations, and significance tests are not affected. A relative sampling weight is a sampling weight divided by the mean of the sampling weights and is denoted as:

$$w_{rel_{ij}} = \frac{w_j}{\sum_j \frac{w_j}{n}} \qquad (2.12)$$

Standard Error Estimation with CSSD

There are several techniques for estimating standard errors that take into account the sampling design. Of these, three were chosen as candidates for simulation in the current study. The first approach, the design effect adjustment is a post hoc adjustment to normally calculated standard errors. The remaining two (the jackknife and bootstrap) are resampling approaches that require an iterative process to arrive at standard error estimates.

Design Effect (DEFF) Adjustment

The design effect (DEFF) is a ratio of the design based standard error to the same estimator as if it were from a SRS of the same size. For a sample proportion estimator, $\hat{P}$, based on *n* observations it is denoted as:

$$DEFF(\hat{P}) = \frac{V(\hat{P})}{\left(1-\frac{n}{N}\right)\frac{P(1-P)}{n}}. \qquad (2.13)$$

If all stratum proportions are equal, the DEFF under stratified sampling will typically be less than one (Lohr, 1999). This is due to the increase in precision associated with stratification noted earlier.

In the case of cluster sampling, the ICC, ρ, is typically positive thus leading to a loss of precision and a DEFF greater than one. The DEFF for a one-stage cluster sample with PSUs of equal size, *M*, is approximately:

$$DEFF = 1 + (M-1)\rho \qquad (2.14)$$

(Kish, 1965; Lohr, 1999). Thus, the DEFF is dependent on both the ICC and *M*. In a design involving both stratification and clustering, it is not possible to know beforehand whether the DEFF will be greater than or less than one. This will depend on whether more precision was gained by stratification or lost by clustering.

Kish (1965) suggested that an estimate of the effective sample size, *n'*, be utilized in the calculation of standard errors instead of *n*:

$$n' = \frac{n}{DEFF}. \qquad (2.15)$$

More often, estimated standard errors are adjusted by multiplication by the quantity $\sqrt{DEFF}$.

Jackknife

The jackknife approach was proposed by Quenouille (1949; 1956) as a method for reducing bias in statistical estimates. This procedure was later adopted to estimate variances and associated confidence intervals (Tukey, 1958; Mosteller & Tukey, 1968).

The most common approach to estimating jackknife variances is a method known as the delete-one jackknife method. Others forms of the jackknife have been proposed and are explained in detail elsewhere (Shao & Tu, 1995). All subsequent references to the jackknife in the current investigation are intended to refer to the delete-one approach.

In a stratified cluster sample with *H* strata, n_h PSU's are chosen to be sampled from stratum *h.* To create a jackknife replicate, one PSU in stratum *h* is omitted and the remaining PSU's in stratum *h* are reweighted to maintain the estimated population size for that stratum. The weights of the remaining units are inflated by a factor of $n_h/(n_h - 1)$ (Rust & Rao, 1996). The analysis of interest is performed on the reduced sample and the resulting parameters are called "jackknife parameter estimates." This procedure is repeated for the entire sample of PSU's or often on a random sample of PSU's. The variance estimates are then calculated from the jackknife parameter estimates.

A widely utilized jackknife variance estimator for a parameter θ is:

$$\hat{V}_{JK}(\hat{\theta}) = \sum_{h=1}^{H} \frac{n_h - 1}{n_h} \sum_{j=1}^{n_h} (\hat{\theta}_{(hj)} - \hat{\theta})^2 \qquad (2.16)$$

where $\hat{\theta}_{(hj)}$ is the estimator and the same form as $\hat{\theta}$ but with PSU *j* of stratum *h* omitted (Lohr, 1999; Skinner, Holt, & Smith, 1989, p. 53). This has been shown to be a conservative variance estimate, that is, it produces relatively larger variance estimates than other methods.

The number of jackknife replicates depends on the desired level of accuracy and potentially on computational efficiency. Several authors have discussed this choice (e.g. Fay, 1985; Patterson, 1998; Wolther, 1985) without reaching a consensus. However, Fay (1985) and Patterson et al. (2002) suggested 20 – 60 groups are

satisfactory based on Monte Carlo simulation results. Based on this suggestion the current investigation utilized 25 jackknife replicates.

Bootstrap

The bootstrap resampling approach was developed by Efron (1979; 1981) and was extended to complex survey samples by Rust and Rao (1996). In this approach, resampling with replacement simulates the sampling distribution of the parameter of interest. For example, a sample of size n_h is taken from stratum h, with replacement. A weight is calculated for this replicate and all parameters of interest are calculated based on this bootstrap sample. This process is repeated a specified number of times, and the standard deviation of these bootstrap samples provides the bootstrap standard error. An advantage of this approach is the avoidance of the normality assumption (Skinner, Holt, & Smith, 1989, p. 54).

Mixtures of Logistic Regressions with CSSD

As noted above, latent class regression models have been developed for a variety of data types, including logistic regression with binary outcome data (Wedel & DeSarbo, 2002). However, the literature is lacking in the treatment of these models with data from complex sample designs. The current investigation incorporated sampling weights by means of the pseudo-likelihood approach and applied this technique to the estimation of latent class regression models. Previous research on latent class regression has shown that the Newton-Raphson procedure can be used to maximize the likelihood across the entire parameter space (Lwin & Martin, 1989). Monte Carlo simulation results support the performance of this algorithm under a variety of data conditions (Wedel & DeSarbo, 1992).

Since LCLR models have not been rigorously developed for use with complex sample survey data, methods for estimating variances are not reported in the literature. While software is available to perform these calculations for standard logistic regression (e.g. STATA, SUDAAN), no widely available software yields estimates of standard errors for LCLR.

CHAPTER III
METHODS

Research Design

The purposes of this study were to (1) develop and test methods for incorporating complex sample designs in LCLR and (2) to evaluate several competing techniques for estimating variances for LCLR parameters (i.e. the bootstrap, jackknife, DEFF adjustment). To address the first of these, the LCLR, likelihood function, and associated SAS programs were written to incorporate complex sample designs. Simulated parameter estimates incorporating stratification, clustering, and weights were compared to empirically derived estimates. These survey-adjusted estimates were compared to results from simulations where the sampling design is ignored. The second purpose was addressed by simulation where bootstrapped, jackknifed and DEFF adjusted standard errors were compared with estimated "true" variances

Observations were simulated to represent a large, simulated population of size *N*, with *K* PSU's each of equal size, *M*. The sample was designed with two strata of equal size to ensure a simple, easy to understand design, though large surveys rarely consist of only two strata. For example, the Indiana Youth Tobacco Survey (IYTS) sampled two PSU's from each of 21 strata for a total of 42 sampled PSU's. Additionally, the IYTS incorporated over-sampling of Hispanics, to allow for meaningful subgroup analyses. As stated, the current investigation simulated two strata of equal size. However, one was "weighted down in size" to reflect an over-sampling of one strata.

Fixed Factors

The sample size, number of latent classes and the number of coefficients estimated were fixed. A sample size of 3000 was chosen, with two strata of equal size, $n_1 = n_2 = 1500$. This sample size was chosen because it is similar to the sample size of the IYTS ($n_{IYTS} = 2860$). Within each stratum, each PSU was of equal size, $m = 20$. This size was chosen because it is large enough to demonstrate a clustering effect (Patterson, 1998; 2002). It was assumed that each of the two strata and the two strata combined could be partitioned into two latent classes (LC) ($\iota = 1, 2$). Within each stratum and when combined, the proportion in LC1 and LC2 will be denoted as θ_t and $(1 - \theta_t)$ respectively. Finally, nine parameters were estimated. These include the most important, the LC mixing proportion, θ_t, the intercept constants (β_{01}, β_{02}) and six regression coefficients (β_{11}, β_{21}, β_{31}, β_{12}, β_{22}, β_{32}), Thus, the number of variables simulated and used in each model was fixed at four (one binary outcome and three independent variables), resulting in the following model

$$q = \theta[\beta_{01} + \beta_{11} + \beta_{21} + \beta_{31}] + (1-\theta)[\beta_{02} + \beta_{12} + \beta_{22} + \beta_{32}]. \qquad (3.1)$$

Preliminary analyses on the IYTS data identified candidate parameters to simulate ensuring realistic disparity between latent classes. Table 1 displays the regression coefficients that were simulated.

Table 1: Simulated Regression Coefficients

Latent Class 1	Latent Class 2
$\beta_{01} = 6.8$	$\beta_{02} = -2.5$
$\beta_{11} = -2.5$	$\beta_{12} = -0.2$
$\beta_{21} = -0.3$	$\beta_{22} = 0.5$
$\beta_{31} = 0.2$	$\beta_{32} = 1.2$

Manipulated Factors

The following aspects of the simulation were manipulated: the mixing proportion parameter, θ (3), clustering (2), and the sampling weights (2). The study investigated 3 scenarios for θ, and two for each of the sample aspects yielding a 3 x 2 x 2 design, with twelve cells of interest. Within each of these cells, parameter estimates and standard errors estimated under differing conditions were compared. Data for each of these scenarios were simulated and analyzed. A conceptual representation of the study is presented in Table 2.

In latent class analyses in general, the parameter that is of most interest is the mixing proportion, θ. For the current investigation, stratum-specific θ 's were defined as follows: θ_1 is the proportion in LC1 in the first stratum and θ_2 is the proportion in LC1 in the second stratum. The population mixing parameter, θ_{pop}, is a function of θ_1 and θ_2 and is denoted as:

$$\theta_{pop} = p\theta_1 + (1-p)\theta_2 \qquad (3.2)$$

where p is the proportion of the total population in stratum 1. In the current investigation $p = 0.8$, due to the fact that stratum 2 was weighted down, as will be discussed in detail.

For the simulations, θ_1 and θ_2 were randomly generated for each cluster using a β distribution in a manner previously described by Patterson (1998, 2002). This approach allows for the introduction of intra-class correlation to the simulated clusters. The β distribution has values that lie in the [0,1] interval making it useful for generating probabilities and proportions. The distribution is identified by two parameters, ν and ω that can be selected such that the mean is located at a desired point along the [0, 1] axis.

Table 2
Simulation Design Specifications
Number of strata = 2, 75 Sampled PSU's of size 20 per stratum

Θ_{pop}	*0.8*	*0.66*	*0.48*
Θ_1 Θ_2	**0.9 0.4**	**0.7 0.5**	**0.5 0.4**
ICC = .01	β(90, 10) β(40, 60)	β(70, 30) β(50, 50)	β(50, 50) β(40, 60)
	SRS	SRS	SRS
	Bootstrap – 100 reps	Bootstrap – 100 reps	Bootstrap – 100 reps
	Jackknife – 25 reps	Jackknife – 25 reps	Jackknife – 25 reps
	$DEFF_{jack}$	$DEFF_{jack}$	$DEFF_{jack}$
	$DEFF_{boot}$	$DEFF_{boot}$	$DEFF_{boot}$
ICC = .1	β(9, 1) β(4, 6)	β(7, 3) β(5, 5)	β(5, 5) β(4, 6)
	SRS	SRS	SRS
	Bootstrap – 100 reps	Bootstrap – 100 reps	Bootstrap –100 reps
	Jackknife – 25 reps	Jackknife – 25 reps	Jackknife – 25 reps
	$DEFF_{jack}$	$DEFF_{jack}$	$DEFF_{jack}$
	$DEFF_{boot}$	$DEFF_{boot}$	$DEFF_{boot}$

Note: The above design was applied to both weighted and unweighted analyses

The β distribution is symmetric about 0.5 when ν = ω. Additionally, the β distribution is skewed to the right when ν < ω and to the left when ν > ω. The probability density function of the β(ν, ω) distribution is:

$$f\left(\theta \mid \nu, \omega\right) = \frac{\Gamma(\nu+\omega)}{\Gamma(\nu)\Gamma(\omega)}\theta^{\nu-1}(1-\theta)^{\omega-1} \qquad (3.3)$$

with mean $\frac{\nu}{\nu+\omega}$ and standard deviation $\sqrt{\frac{\nu\omega}{(\nu+\omega)^2(\nu+\omega+1)}}$. The ICC for this distribution is $(\nu+\omega+1)^{-1}$ (Brier, 1980; Patterson, 1998). The β distributions that were used are summarized in Table 3.

The β distributions were selected to simulate three levels of θ_{pop}. The simulations investigated population mixing parameters that varied from moderate (θ_{pop} =0.48) to more extreme values (θ_{pop}=0.66 and θ_{pop} =0.80). Values of θ were selected such that there will be a varying amount of disparity between strata. Pilot study on the IYTS indicated that θ_{pop} =0.80 for some of the models investigated. Table 4 displays the values of θ_{pop} and the values of θ_t for each stratum.

As noted above, clusters of size m = 20 were generated. Values of θ were randomly generated from one of the β distributions previously described, thus introducing the intra-class correlation. That is, observations within a given cluster were more alike than observations from other clusters. This was done by using the randomly generated value of θ to generate the 20 observations for each cluster. The process was repeated 75 times for each stratum, thus yielding $n_1 = n_2 = 1500$. The result was cluster level variation of the proportion in LC1. This approach has been previously

Table 3
β Distribution Summary

ν	ω	Mean	Standard Deviation	ICC (ρ) $(\nu+\omega+1)^{-1}$	DEFF $1+(M-1)\rho$
10	90	0.1	0.029	0.01	1.82
30	70	0.3	0.046	0.01	1.82
40	60	0.4	0.049	0.01	1.82
50	50	0.5	0.049	0.01	1.82
1	9	0.1	0.091	0.10	2.73
3	7	0.3	0.131	0.10	2.73
4	6	0.4	0.148	0.10	2.73
5	5	0.5	0.151	0.10	2.73

Table 4
Strata and Population Values of θ

Stratum 1 θ_1	Stratum 2 θ_2	Population θ_{pop} $p\theta_1 + (1-p)\theta_2$
.9	.4	0.80
.7	.5	0.66
.5	.4	0.48

Note: p = 0.8

reported and verified through simulation (Patterson, 1998, 2002). Two levels of ICC were investigated, 0.01 and 0.1. These values were chosen as they are reflective of the ICC that has been reported for various items on the IYTS and are similar to those reported in other large surveys. As noted in Table 1, with cluster sizes of m_i=20, these ICCs of 0.01 and 0.1 result in DEFF = 1.18 and 2.73, respectively.

The final manipulated design feature was sample weighting. As previously stated, the IYTS uses over-sampling of certain subgroups. The current investigation also incorporated simulated over-sampling. Two scenarios of weighting were investigated, one ignoring and one incorporating the weights. In the first scenario, data were simulated for 1500 observations in each stratum, the models of interest fit to the data, and standard errors estimated. In the second, weighted scenario, 1500 observations were also simulated in each stratum. However, to incorporate "over-sampling" the second stratum was weighted down to a sample size of 375. Thus, in the weighted scenario, the effective, weighted sample size was 1875. The weighted sample sizes were $n_{w1} = 1500, n_{w2} = 375, N_w = 1875$, for weighted stratum one, weighted stratum two, and weighted combined sample respectively. Thus the proportion of observations in the first stratum is 0.8. Ratios of this magnitude (4:1) are common in complex survey designs and are large enough to allow for the investigation of bias introduced by ignoring weights. Therefore, two sets of simulations were done, one incorporating the weights and one ignoring them. The current investigation was weighting down to account for over-sampling and not weighting up to a population, so relative weights were not appropriate. The weights for stratum one and stratum two were 1.0 and 0.25 respectively.

Variance Estimation

This study investigated the performance of several competing variance estimation techniques. The estimated variances were compared to "true" variances where true variances were empirically derived based on 10,000 replicates in a manner similar to Patterson et al. (2002). The first variance estimation technique would be appropriate if SRS had been used to select the sample. It was expected that this method would yield variances that are consistently liberal (i.e. too small).

Simulated variances for each experimental condition were estimated first, via the jackknife. A jackknife replicate was generated by sequentially leaving out a single PSU, reweighting the data, and then estimating the LCLR parameters. The jackknife weights (w_h) for strata h=1 and h =2 were calculated as follows:

$$w_1 = \frac{n_{hw}}{n_{hw-1}} = \frac{1500}{1480} = 1.0135, \; w_2 = \frac{n_{hw}}{n_{hw-1}} \frac{375}{1480} = 0.25338 \qquad (3.4a, 3.4b)$$

The mean of the jackknife parameter estimates and their variances were calculated. As previously indicated, 25 jackknife replicates have been shown to be adequate (Fay, 1985) and this number was performed to generate the simulated variances.

Variances were also estimated via the bootstrap. Each replicate was created by randomly sampling 75 PSU's from each stratum with replacement. Efron and Tibshirani (1993, p. 52) state that for estimating standard errors, more than 100 replications are typically unnecessary. Therefore, bootstrap standard errors were the standard deviation of a simulated sampling distribution based on 100 replications.

The final variance estimation technique that was studied was the DEFF adjustment. The DEFF's were calculated in two ways. The first was the ratio of the

jackknife variance estimate to the SRS variance estimate and the second was the ratio (for each parameter) of the bootstrap variance to the SRS variance and are denoted as:

$$DEFF_{jack} = \frac{V_{jack}}{V_{SRS}} \text{ and } DEFF_{boot} = \frac{V_{boot}}{V_{SRS}} \text{ respectively.} \qquad (3.5a, 3.5b)$$

The square roots of DEFF's were then calculated and multiplied by the SRS standard error as the adjustments. That is, $se_{DEFF} = \sqrt{DEFF} * se_{SRS}$.

Simulation Details

The simulation code was written in SAS version 8.2 (SAS Institute, 2001). The Newton-Raphson method was used to maximize the non-linear likelihood functions. Previous studies have indicated that this method is suitable for this type of analysis (Wedel & DeSarbo, 1992). Proc NLP was the SAS procedure that was used for optimization. The NLP procedure is available in the operation research (OR) component of SAS. This procedure uses the gradient and Hessian matrix and thus requires that the objective function have continuous first and second-order derivatives. The algorithm uses a pure Newton step when the Hessian is positive definite, otherwise a combination of ridging and line-search is done to compute successful steps. If the Hessian is not positive definite, a multiple of the identity matrix is added to the Hessian to make it positive definite (SAS Institute, 1999).

The method proceeded in the following manner. (1) Data were randomly generated and the model of interest was fit yielding parameter estimates and standard errors. These were saved and will be referred to as the SRS estimates. (2) Jackknife replicates were taken from these simulated data and the resulting jackknife estimates were saved. (3) Bootstrap samples were taken from these same data and estimates were

saved. The above design resulted in three sets of parameter estimates (SRS, jackknife, and bootstrap) and five sets of variance estimates (SRS, jackknife, bootstrap, $DEFF_{jack}$, $DEFF_{boot}$). Each simulation consisted of 1,000 replications with a convergence criterion set to 10^{-5}. Finally, the maximum number of iterations for convergence was set to 500.

Cases with solutions that did not converge in the specified number of iterations were noted and handled as follows. A boundary can occur in the initial estimation, during the jackknife, or during the bootstrap. If it occurred during the initial estimation, the parameter was set to zero and its variance was not estimated. If this occurs during the jackknife or bootstrap, estimates for that replicate were set to zero.

IYTS and Simulation Origin

The IYTS was the first survey designed to provide comprehensive, baseline data on the prevalence of tobacco use among Indiana youth and was designated to be used to guide and evaluate youth tobacco-use prevention programs. It is the first survey to provide Indiana with data that can be compared to other states or the national average. The survey also measured knowledge and attitudes about tobacco, the impact of media and advertising, minors' access to tobacco, tobacco-related school curricula, exposure to environmental tobacco smoke, and cessation of tobacco use. The survey and sampling method was developed by the Centers for Disease Control and Prevention (CDC) to be a scientifically valid, random sampling of Indiana's youth. All public high schools containing grades 9, 10, 11, or 12 and all public middle schools containing grades 6, 7, or 8 were included in the sampling frame. A two-stage cluster sample design was used to produce a representative sample of students in grades 6 through 12.

LCLR provides a novel approach to investigate unobserved subgroups of youth that differ in terms of what effects their susceptibility to experiment with or initiate tobacco use. Specifically, there is a dichotomous survey item that asks if a respondent thinks that he or she will try a cigarette in the near future. The covariates that were identified as differing between latent class regressions were age and two Likert-type attitude items. As noted above, some pilot analyses with these data dictated the design of the simulations to ensure that they mimicked real world data (see Table 1).

Statistical Analysis

The analyses focused on parameter bias, variance estimates, and confidence interval coverage. Relative bias was calculated as the ratio of the simulated parameter values to the true values that were calculated beforehand. Unadjusted (SRS), jackknife, bootstrap, and DEFF adjusted variances were compared to the derived "true" variances. The "true" variances were determined for each condition using the following equation for each parameter:

$$V_{true} = \frac{1}{10{,}000} \sum_{i=1}^{10{,}000} (\hat{\theta}_i - \theta)^2 \qquad (3.6)$$

Parameter estimates, unadjusted (SRS) , jackknife, bootstrap, and DEFF adjusted variances were assessed in terms of their relative bias:

$$Bias_{rel} = \frac{\sum_{i=1}^{1000} \frac{\theta_i}{1000}}{\theta_{true}} \qquad (3.7)$$

where θ_i is the parameter estimate from the I^{th} simulation. Additionally, 95% confidence intervals for parameter estimates were calculated as:

$$C.I. = \hat{\theta} \pm 1.96\sqrt{v_i} \qquad (3.8)$$

These intervals were utilized to investigate the interval coverage of the “true” parameters.

CHAPTER IV
RESULTS

Parameter estimates and bias are discussed followed by a discussion of the variance estimators in terms of bias and confidence interval coverage. Then, the results for the example based on the Indiana Youth Tobacco Survey based on the "best" method are presented.

Parameter Estimates and Bias

The parameter estimates from the various simulations and associated biases are presented in Appendix A where tables are appropriately prefixed. Each of the estimated parameters is presented in the far left column, the estimated true values in the second column, and the estimates from the various conditions are presented as denoted by the other column headings.

The weighted estimates presented are nonlinear and obtained via pseudo-maximum likelihood. Thus, they are consistent but may be biased (Roberts, Rao, and Kumar, 1987). As noted above, in latent class analysis it is the mixing proportion θ_{pop} that is of the most interest. Additionally, the results for the regression coefficients were consistently similar to those for θ_{pop}. Thus, for the sake of simplicity and brevity, the following discussion of results are restricted to θ_{pop}. Complete results of all parameters estimates are presented in the appendix.

When weights were used in estimation, the resulting values of θ_{pop} were quite similar to the estimated true values (Tables A1 through A6) with no weighted estimate of θ_{pop} showing more than 4.96% bias. When the weights were ignored, bias was found

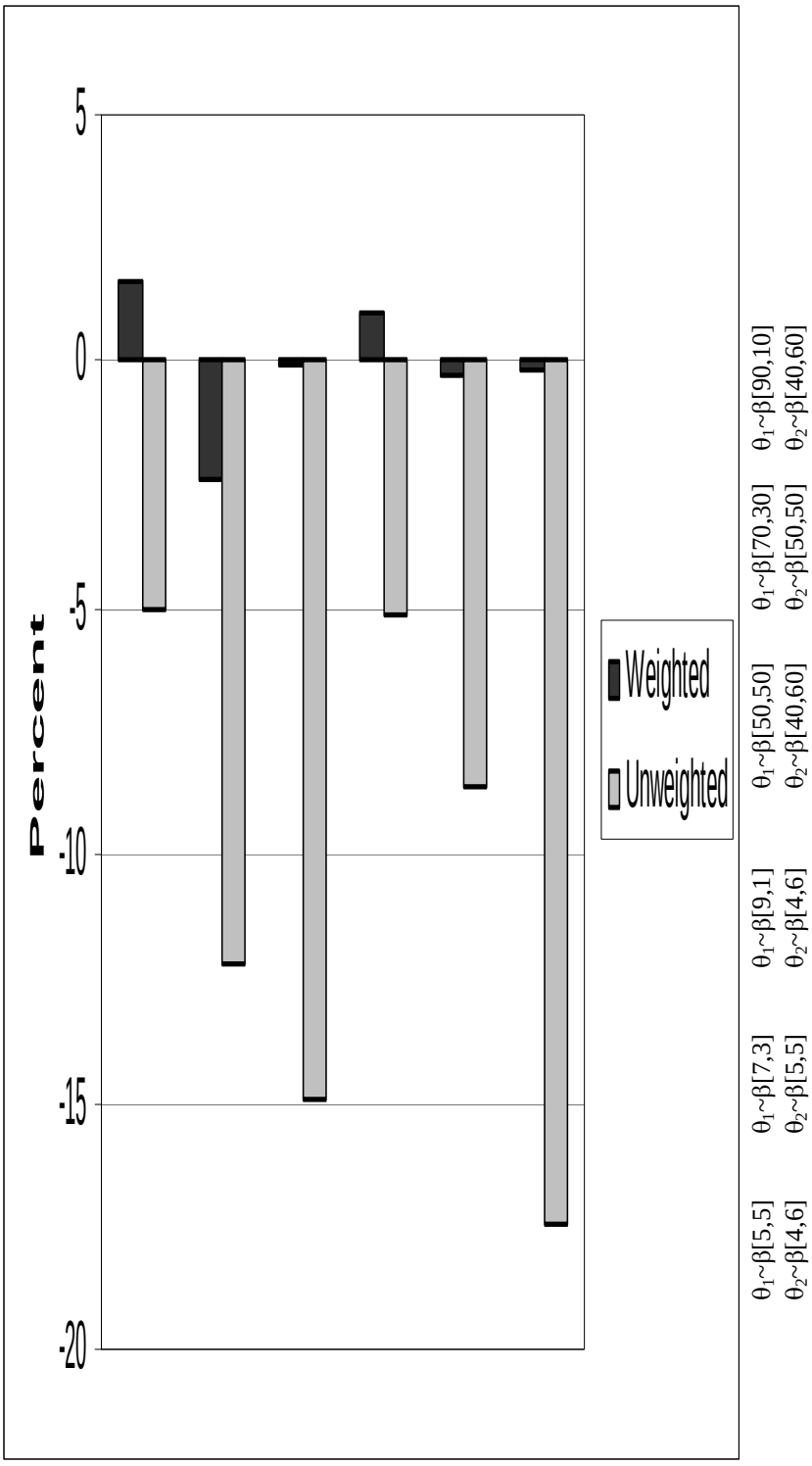

Figure 1: Relative Bias of SRS Estimates of θ_{pop}

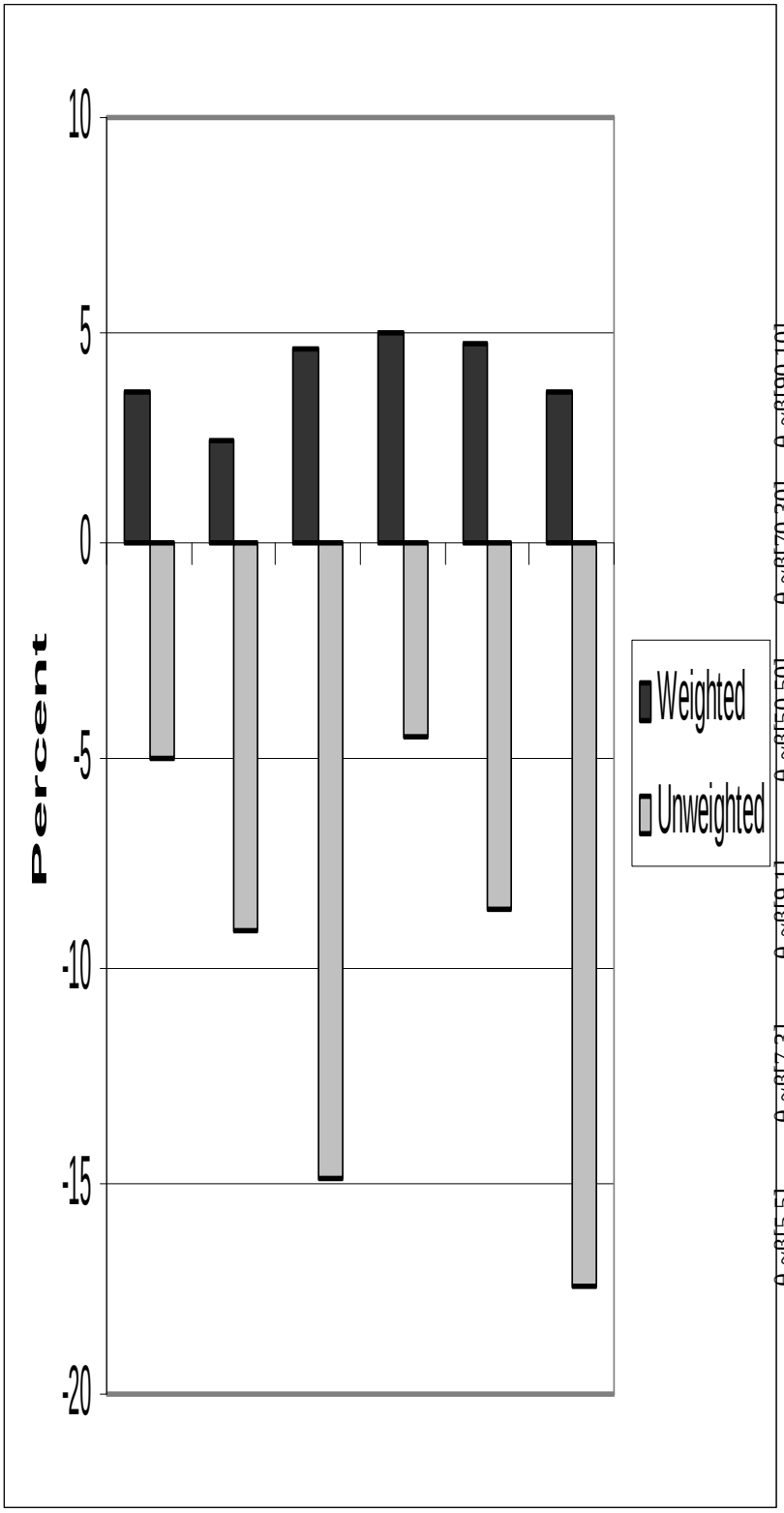

Figure 2: Relative Bias of Jackknife Estimates of θ_{pop}

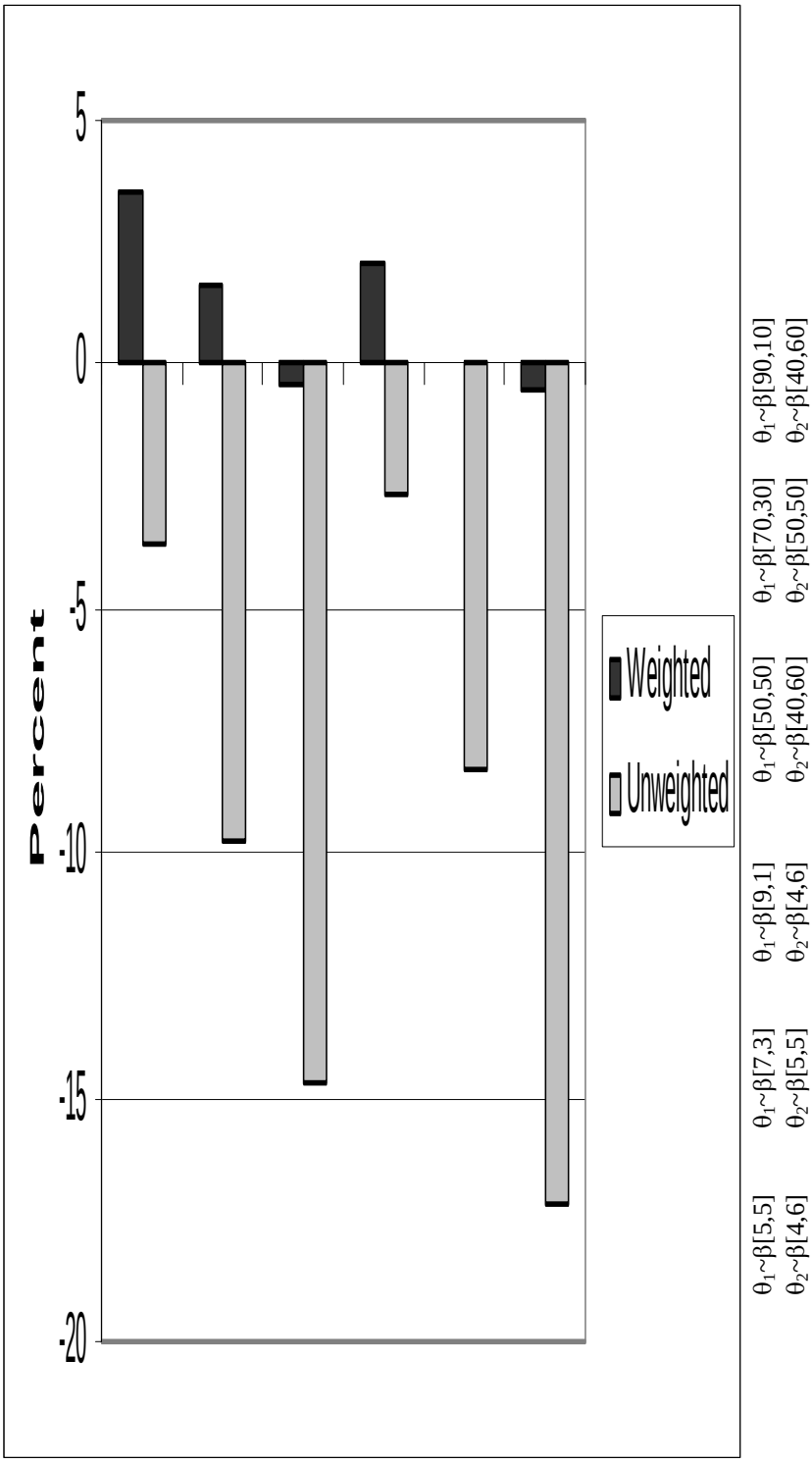

Figure 3: Relative Bias of Bootstrap Estimates of θ_{pop}

to be much larger, exceeding 17% in several cases. Figures 1 through 3 depict the amount of parameter bias detected in each of the three parameter estimate scenarios.

In general, the amount of bias increased as the value of θ_{pop} became more extreme (i.e. deviated from .5 closer to 1). Averaging over results, it is apparent that the weighted SRS (mean bias -0.04%) was less biased than the weighted bootstrap (mean bias 1.01%) which was less biased than the weighted jackknife (mean bias 3.97%) estimates. Among the unweighted results, there does not appear to be a clearly superior method in relation to parameter bias. That is, all three unweighted methods yielded comparably biased estimates.

Variances

As noted above, the estimated "true" variances were calculated as the sample variance of 10,000 replications of the model of interest fit to simulated data. Resulting standard error estimates were evaluated first in terms of the ratio of a given estimate to the derived true estimate. A complete summary of all standard error estimates can be found in Appendix A (Tables A7 through A12).

As hypothesized, standard error estimates ignoring the complex sample design and assuming SRS consistently resulted in standard errors of θ_{pop} that were smaller than the true values. These are presented in Figure 4. These standard errors underestimated the true values by values of 0.45 to 18.66%. Thus, inferential tests based on these estimates may be undesirably liberal. Jackknife estimates overestimated standard errors of θ_{pop} in 66.66% of the scenarios by values of 0.66 to 13.25%. Bootstrap estimates overestimated the true standard errors of θ_{pop} in 83.33% of the scenarios investigated by values of 2.17 to 21.95%. Thus, resulting tests would be more conservative.

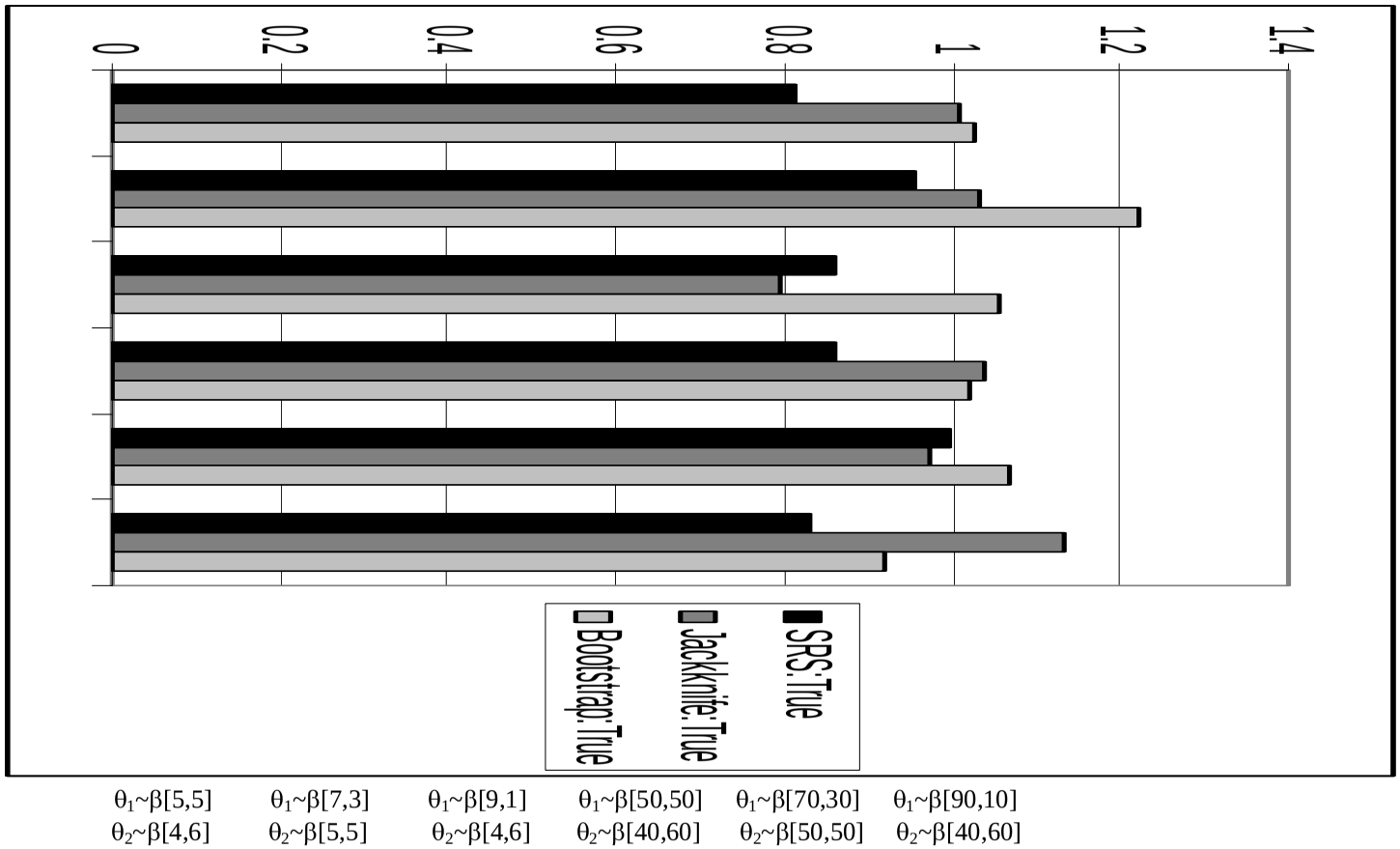

Figure 4: Ratios of Variance Estimates to "True" Variances of θ_{pop}

The impact of clustering was investigated via design effects (DEFF's). As stated in Chapter 3, two design effects were calculated for each parameter based on the jackknife and bootstrap standard errors. The design effects are presented for each parameter in Appendix A (Tables A7 through A12). As expected, the scenarios utilizing the values from the β distribution intended to result in larger ICCs (i.e. $\theta_1 \sim \beta[5,5]$ $\theta_2 \sim \beta[4,6]$; $\theta_1 \sim \beta[7,3]$ $\theta_2 \sim \beta[5,5]$; $\theta_1 \sim \beta[9,1]$ $\theta_2 \sim \beta[4,6]$) did result in larger DEFF's than the scenarios intended to exhibit weaker clustering effects. However, the design effects were consistently smaller than those suggested by equation (2.14). For example, the observed design effects for θ_{pop} from $\theta_1 \sim \beta[5,5]$ $\theta_2 \sim \beta[4,6]$ were 1.258 and 1.237 for the bootstrap and jackknife variances respectively but the expected value for each was 2.73. Similarly, the observed design effect for θ_{pop} from $\theta_1 \sim \beta[50,50]$ $\theta_2 \sim \beta[40,60]$ were 1.184 and 1.206 for the bootstrap and jackknife variances respectively but the expected value for each was 1.82.

Confidence Interval Coverage

Standard error estimates were investigated in terms of coverage where coverage was defined as the percentage of replicates in which a 95% confidence interval constructed by equation (3.8) included the true parameter value. A complete list of coverage for every parameter can be found in Tables A13 through A18.

For θ_{pop},, the unweighted analyses resulted in coverage ranging from a minimum of 75.00%, occurring in the $DEFF_{boot}$ condition to a maximum coverage of 89.33% occurring in the bootstrap condition. Additionally, the coverage observed in the unweighted conditions varied as a function of both the clustering effect and the

difference between θ_1 and θ_2. That is, coverage was greatest in the condition with $\theta_1 \sim \beta[50,50]$ $\theta_2 \sim \beta[40,60]$ and poorest in the condition with $\theta_1 \sim \beta[90,10]$ $\theta_2 \sim \beta[40,60]$.

The coverage for each of the weighted scenarios are presented in Figure 5. When weights were incorporated into the analyses, the coverage improved considerably with coverage ranging from 85.1% to 95.3%. Of the methods that were compared, the bootstrap consistently resulted in the best coverage. In fact, in every weighted scenario except one, the bootstrap coverage was largest. The DEFF adjustment to SRS standard errors resulted in coverage that in some scenarios outperformed the jackknife. However, other than the bootstrap appearing to be clearly the best choice no other systematic pattern emerged.

Example Based on IYTS

The IYTS is a unique survey designed to allow for the quantification of various Indiana youth attitudes, beliefs, and behaviors towards tobacco use. Pilot investigation was undertaken to identify a candidate model that was used as the basis of the larger simulation study already presented.

The dichotomous dependent variable that was used was the response to the survey item "Do you think you'll try a cigarette anytime soon?" Possible responses were either "no" or "yes/already tried smoking." This item was intended to capture intent to initiate tobacco use/abuse. The first of the covariates (X1) was a four category Likert item that asked "If one of your friends offered you a cigarette, would you smoke it?" Possible responses ranged from "Definitely Yes" which was coded as a 1 to

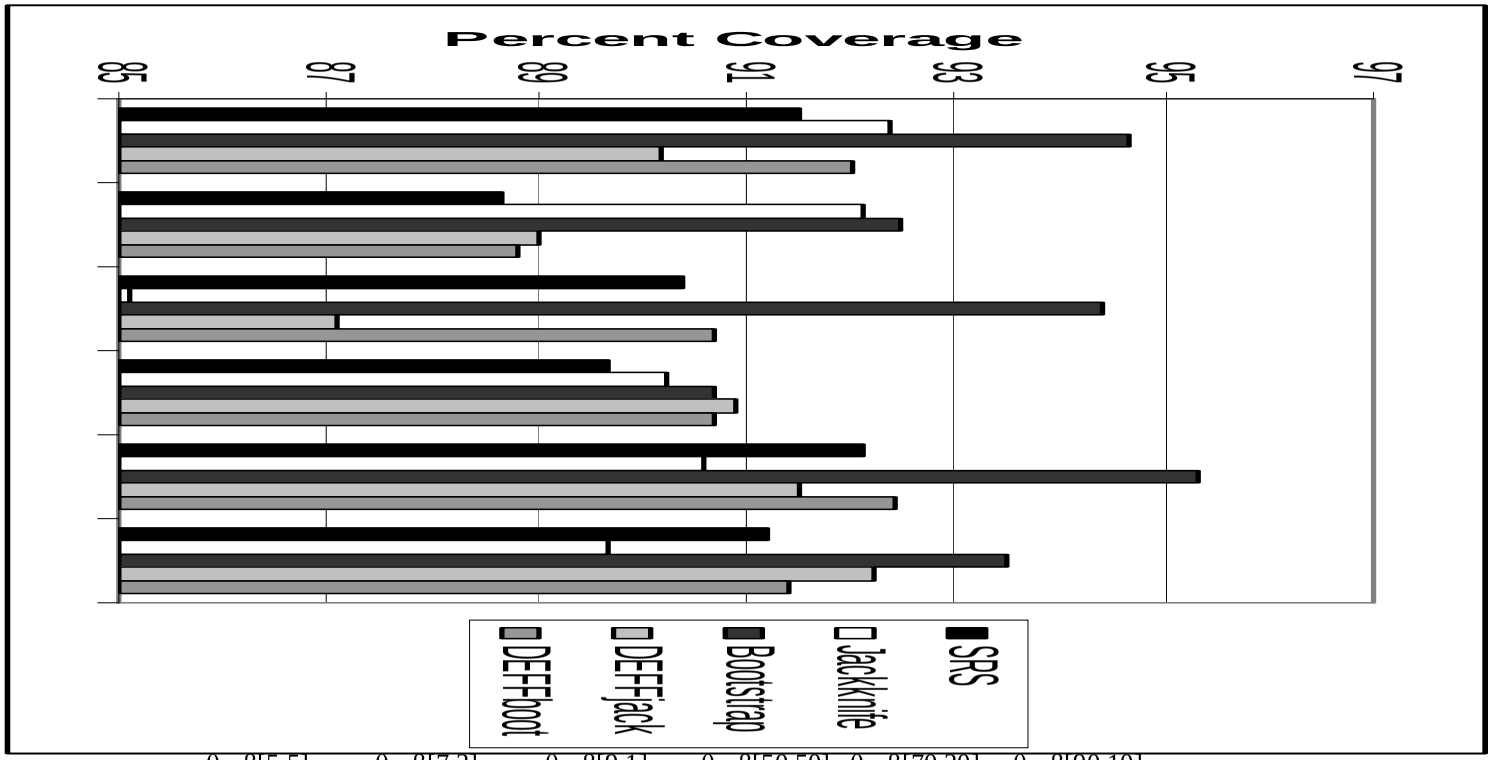

Figure 5: Confidence Interval Coverage for Weighted Estimates of θ_{pop}

"Definitely Not" coded as a 4 with the middle responses of "Probably Yes" and "Probably Not" were coded as 2 and 3 respectively. This item was intended to capture susceptibility to peer influence regarding tobacco. The second covariate (X2) was a Likert item that asked if respondents believed that "People get addicted to tobacco just like cocaine or heroin." Again, responses ranged from 1 ("Definitely Yes") to 4 (Definitely No"). The final covariate (X3) was an ordinal variable representing age that was coded as one for age twelve or younger up to eight for those aged nineteen or older.

Initially, a weighted, one class LR and a three class LCLR model were fit to these data. The resulting one class model is contained in Table 5. The one class model was then compared to the two and three class alternative to determine which fit the

Table 5: Weighted IYTS LR Coefficients

Covariate	*Estimate*
Constant	$\beta_0 = 3.97$
Peer Influence	$\beta_1 = -1.63$
Addiction	$\beta_2 = -0.01$
Age	$\beta_3 = 0.25$

IYTS data better. Akaike (1974, 1981) proposed the Akaike's information criteria (AIC) as a way to choose the best fitting among competing models. AIC is a statistic commonly used to compare latent class models. For a model consisting of g parameter estimates AIC is denoted as:

$$AIC = -2LNL + 2g \qquad (4.1)$$

The model with the minimum AIC is preferred. Table 6 contains the values of LNL and AIC for the one, two, and three class models. The two class LCLR model was chosen, indicating that the population in question was a heterogeneous one. That is, these data

are better represented by a mixture of two LRs than one LR or a mixture of three. Other model selection techniques such as the Bayesian information criteria (BIC) have been suggested by Schwarz (1978), Box, Jenkins, and Reinsel (1994) among others. BIC is of the form:

$$BIC = -2LNL + \ln(n)g \qquad (4.2)$$

BIC was not utilized in the current investigation because it was unclear what quantity to utilize for the sample size *n*. The choice is between the total sample size and the number of PSUs. In order to avoid the nebulous nature of this choice, AIC was chosen over BIC for model selection.

Table 6: Log Likelihood and AIC

Model	*LNL*	*AIC*
One Class	-212709.43	425426.86
Two Class	-210185.03	420388.07
Three Class	-210181.48	420390.96

Table 7 displays the regression coefficients resulting from the weighted two class LCLR model fit to the 2001 IYTS data. The mixing proportion for this model was $\theta = 0.82$. Latent class one comprises 82% of the population. This first class appears to be more susceptible to peer influence in that those that indicated that they would

Table 7: Weighted IYTS LCLR Regression Coefficients

Covariate	*Latent Class 1*	*Latent Class 2*
Constant	$\beta_{01} = 6.81$	$\beta_{02} = -2.49$
Peer Influence	$\beta_{11} = -2.47$	$\beta_{12} = -0.14$
Addiction	$\beta_{21} = -0.33$	$B_{22} = 0.51$
Age	$\beta_{31} = 0.15$	$B_{32} = 1.21$

Note: $\theta = 0.82$

accept a friend's offering of a cigarette were more likely to indicate that they either intended to try or have already tried a cigarette. In both classes it appears that those that are older are more likely to indicate that they would smoke. The classes differ in their knowledge of the addictive nature of tobacco, however these coefficients are not of significant magnitude. One interpretation would be to classify these groups as "susceptibility classes." That is, 82% of Indiana youth tend to be more susceptible to the influence of their peer in terms of tobacco use. The other 18% tend to not be affected by their peers, but become more likely to smoke as they age. Additional variables would need to be investigated to further differentiate these latent subgroups. Indiana has one of the highest youth smoking rates in the United States, which was one of the motivations for fielding this survey. Based on these findings, future counter marketing or public service announcements in Indiana should address the peer pressure issue.

Based on the simulation results, it was determined that relying on weighted parameter estimates would result in less bias and bootstrapped standard errors would produce the most conservative variance estimates. However, unweighted parameter estimates and standard errors were investigated as well for comparative purposes.

Table 6 displays the unweighted parameter estimates, a complete summary of parameters and standard errors can be found in Table A19. Ignoring the weights resulted in biased parameters, even resulting in the change of sign for β_{21} in latent class one. This indicates again, that weights must be included for results to be appropriate. Most importantly, the unweighted estimate of θ was quite biased.

Table 8: Unweighted IYTS LCLR Regression Coefficients

Latent Class 1	*Latent Class 2*
$\beta_{01} = 6.12$	$\beta_{02} = -3.46$
$\beta_{11} = -1.59$	$\beta_{12} = -0.20$
$\beta_{21} = 0.22$	$B_{22} = 0.63$
$\beta_{31} = 0.06$	$B_{32} = 2.00$

Note: $\theta = 0.62$

The IYTS was the result of a sampling design that was quite different from the simplified scenarios used in the simulations. In addition, the IYTS utilized much more extreme weights for several subgroups. These two factors no doubt influenced the amount of observed bias. However, the bootstrap approach once again produced standard errors that consistently over-estimated their SRS analogs.

CHAPTER V
CONCLUSIONS AND DISCUSSION

This investigation was primarily concerned with two areas. First, to investigate potential bias introduced to parameter estimates when sampling weights are not incorporated into the analysis. The second was to investigate competing strategies of variance estimation in the face of varying amounts of intra-cluster correlation introduced by complex survey data.

Parameter Estimates

Typically, when data are analyzed, parameter values obtained from analysis (i.e. not from a replication methodology) are used. The results of the current investigation are in agreement with this practice. However, when analyzing complex survey data, valid parameters can only be obtained if weights are incorporated into the analysis. Additionally, jackknife and bootstrap parameter estimates were more biased than their SRS counterparts when using weights. While this bias was not in all cases extreme, the SRS parameter estimates tended to be less biased. The parameter estimates obtained when weights were ignored were consistently biased. This held true for all estimates SRS, bootstrap, and jackknife.

These findings underscore the importance of incorporating weights. When weights were ignored, the estimates of θ_{pop} were quite biased, in come cases over 17%. In the IYTS example, ignoring the weights resulted in an estimate of θ_{pop} that was biased by 24%.

Variance Estimates

As hypothesized, standard error estimates based on the assumption of SRS consistently under estimated the true values by 0.4 to 18.6%. This approach should be abandoned due to the overly liberal variance estimates it produces. The replication based estimates (i.e. jackknife and bootstrap) produced standard error estimates that consistently over-estimated the true variances.

The jackknife standard errors resulted in estimates that were in general, positively biased. That is, this approach tended to over-estimate true standard errors by .06 to 13%. However, this was not the case in every condition. Specifically, in two of six weighted scenarios the jackknife estimates underestimated the true variance by 3 to 20%. This underestimation can lead to spurious statistical tests. However, the jackknife as implemented here, consisted of only 25 replicates. The jackknife is easy to implement in statistical packages such as SAS. It would be beneficial for future research to investigate the impact of doubling or even tripling the number of jackknife replicates in an attempt to arrive at a suggested number of replicates that would provide more stable estimates of variability.

The bootstrap standard error estimates were also predominantly conservative. In five of the six weighted scenarios, the bootstrap over-estimated the true standard error by 2.1 to 21.9%. However, in one scenario, it underestimated the standard error by 8%. While this is only slightly better than the jackknife, as will be discussed, the bootstrap consistently outperformed it in terms of coverage. It would be beneficial to investigate the impact of increasing the number of bootstrap replicates as well.

The SRS standard error estimates were quite liberal, and the observed design effects were much smaller than expected. The design effect adjustment resulted in variance estimates that were also consistently smaller than were the true values. This approach could be investigated again in simulated scenarios that exhibited larger design effects, perhaps with values of DEFF larger than two.

Based on the current findings, the final recommendation regarding variance or error estimation must be the bootstrap. However, this recommendation is based on a desire to "err on the side of caution." That is, it is preferable to use variance estimates that tend to be conservative (i.e. large) rather than liberal (i.e. small). This is especially true if rejecting the null hypothesis has severe consequences. Additional research could provide a more concrete guide.

Coverage and Recommendations

Confidence interval coverage was greatest when bootstrap standard errors were used to construct the confidence intervals. In fact, the bootstrap was the only method that resulted in 90% coverage or better for θ_{pop} in every weighted scenario. Again, these results may be bolstered by an increase in the number of bootstrap replicates.

Based on the findings for bias, variance estimation, and coverage, it is clear that the preferred method is to fit the model and obtain parameter estimates then approximate variance or standard errors through the bootstrap method. The smallest parameter bias was observed with SRS (non-replicate) estimates, and the most conservative error estimates were obtained via the bootstrap.

Implications for Future Research

1. The number of replicates for both the bootstrap and jackknife approach should be further investigated. An increase in either might significantly decrease the observed bias in both parameter and variance estimate bias. Patterson (1998, 2000) indicated that sixty jackknife replicates resulted in acceptable coverage and minimal bias. Based on the current investigation, the number of jackknife or bootstrap replicates should be increased, possibly doubled.

2. Systematic variation of the simulated sample design would indicate the performance of these methods in other circumstances. For example, it would be useful to simulate a survey sample consisting on more (possibly many more) than two strata and few cluster per strata. Few real world surveys consist of only two strata.

3. It would be of interest to investigate the current methods both on much larger and much smaller samples.

4. Other variance estimation techniques such as linearization should be investigated.

APPENDIX A: RESULTS TABLES

Table A1: Parameter Estimates and Relative Bias

$\theta_1 \sim \beta[5,5]$; $\theta_2 \sim \beta[4,6]$

Weighted Analyses

Parameter	"True" Parameter	SRS	Relative Bias (%)	Jackknife	Relative Bias (%)	Bootstrap	Relative Bias (%)
β_{01}	6.99	6.911	-1.05	7.03	0.65	7.00	0.14
β_{11}	-2.57	-2.44	-4.96	-2.54	-1.35	-2.57	0.11
β_{21}	-0.39	-0.37	-4.08	-0.40	2.47	-0.39	1.08
β_{31}	0.17	0.16	-5.99	0.17	-2.32	0.17	-0.99
β_{02}	-2.91	-3.09	6.45	-3.08	5.93	-2.85	-1.81
β_{12}	-0.10	-0.09	-4.20	-0.09	-5.03	-0.09	-4.08
β_{22}	0.64	0.64	-0.06	0.69	6.98	0.66	2.85
β_{32}	1.46	1.48	0.91	1.54	5.52	1.49	2.13
θ	0.48	0.49	1.65	0.50	3.64	0.500	3.56

Unweighted Analyses

Parameter	"True" Parameter	SRS	Relative Bias (%)	Jackknife	Relative Bias (%)	Bootstrap	Relative Bias (%)
β_{01}	6.99	6.96	-0.39	8.10	15.92	7.49	7.14
β_{11}	-2.57	-2.61	1.66	-2.94	14.33	-2.84	10.36
β_{21}	-0.39	-0.39	-0.22	-0.49	24.65	-0.48	21.81
β_{31}	0.17	0.20	17.18	0.24	35.56	0.18	6.42
β_{02}	-2.9	-2.79	-4.08	-2.79	-4.00	-3.25	11.75
β_{12}	-0.10	-0.12	18.14	-0.12	18.52	-0.10	3.33
β_{22}	0.64	0.59	-7.61	0.59	-7.36	0.72	12.52
β_{32}	1.46	1.40	-4.10	1.40	-4.00	1.66	13.76
θ	0.48	0.45	-5.03	0.45	-5.10	0.46	-3.71

Table A2: Parameter Estimates and Relative Bias

$\theta_1 \sim \beta[7,3]$; $\theta_2 \sim \beta[5,5]$

Weighted Analyses

Parameter	"True" Parameter	SRS	Relative Bias (%)	Jackknife	Relative Bias (%)	Bootstrap	Relative Bias (%)
β_{01}	7.44	7.21	-3.14	7.58	1.78	7.45	0.06
β_{11}	-2.78	-2.6	-4.15	-2.57	-7.45	-2.67	-3.75
β_{21}	-0.35	-0.36	3.27	-0.36	4.09	-0.36	2.63
β_{31}	0.20	0.19	-4.09	0.20	-1.43	0.20	0.48
β_{02}	-3.18	-3.09	-2.72	-3.09	-2.82	-3.19	0.40
β_{12}	-0.06	-0.06	-3.33	-0.06	1.82	-0.06	0.53
β_{22}	0.69	0.67	-2.39	0.69	1.31	0.70	1.38
β_{32}	1.60	1.53	-4.44	1.68	4.67	1.65	3.17
θ	0.67	0.65	-2.39	0.69	2.45	0.68	1.60

Unweighted Analyses

Parameter	"True" Parameter	SRS	Relative Bias (%)	Jackknife	Relative Bias (%)	Bootstrap	Relative Bias (%)
β_{01}	7.44	7.69	3.27	7.63	2.47	7.61	2.30
β_{11}	-2.78	-2.56	-8.01	-3.05	9.93	-2.99	7.74
β_{21}	-0.35	-0.38	10.49	-0.37	6.77	-0.36	4.18
β_{31}	0.20	0.23	14.83	0.22	9.04	0.22	8.99
β_{02}	-3.18	-2.83	-11.06	-2.88	-9.45	-3.54	11.38
β_{12}	-0.06	-0.07	11.96	-0.07	9.31	-0.07	8.49
β_{22}	0.69	0.66	-3.61	0.63	-8.44	0.62	-9.72
β_{32}	1.60	1.49	-6.82	1.43	-10.92	1.44	-10.40
θ	0.67	0.59	-12.16	0.61	-9.16	0.61	-9.71

Table A3: Parameter Estimates and Relative Bias

$\theta_1 \sim \beta[9,1]$; $\theta_2 \sim \beta[4,6]$

Weighted Analyses

Parameter	"True" Parameter	SRS	Relative Bias (%)	Jackknife	Relative Bias (%)	Bootstrap	Relative Bias (%)
β_{01}	7.63	8.08	5.91	7.86	3.05	7.70	0.87
β_{11}	-2.85	-2.99	4.96	-3.03	6.25	-2.98	4.83
β_{21}	-0.35	-0.36	4.63	-0.34	-3.26	-0.34	-2.42
β_{31}	0.21	0.21	-2.88	0.23	10.99	0.22	4.02
β_{02}	-3.23	-3.17	-1.97	-3.14	-2.83	-3.24	0.26
β_{12}	-0.06	-0.06	-2.54	-0.07	6.18	-0.07	4.73
β_{22}	0.70	0.73	3.96	0.74	5.36	0.71	0.78
β_{32}	1.63	1.71	4.59	1.74	6.81	1.69	3.85
θ	0.79	0.79	-0.05	0.82	4.61	0.78	-0.34

Unweighted Analyses

Parameter	"True" Parameter	SRS	Relative Bias (%)	Jackknife	Relative Bias (%)	Bootstrap	Relative Bias (%)
β_{01}	7.63	7.23	-5.20	7.86	2.99	7.86	3.05
β_{11}	-2.85	-2.70	-5.24	-2.96	3.45	-2.95	3.59
β_{21}	-0.35	-0.33	-5.28	-0.34	-1.13	-0.34	-1.15
β_{31}	0.21	0.18	-14.56	0.24	13.55	0.22	4.46
β_{02}	-3.23	-2.96	-8.52	-3.50	8.43	-3.23	-0.04
β_{12}	-0.06	-0.09	37.84	-0.09	34.66	-0.08	23.15
β_{22}	0.706	0.63	-10.33	0.74	5.74	0.72	3.09
β_{32}	1.63	1.48	-9.51	1.75	7.01	1.69	3.58
θ	0.79	0.67	-14.91	0.82	4.47	0.82	4.48

Table A4: Parameter Estimates and Relative Bias

$\theta_1 \sim \beta[50,50]$; $\theta_2 \sim \beta[40,60]$

Weighted Analyses

Parameter	"True" Parameter	SRS	Relative Bias (%)	Jackknife	Relative Bias (%)	Bootstrap	Relative Bias (%)
β_{01}	6.94	6.71	-3.36	7.29	5.08	7.02	1.16
β_{11}	-2.56	-2.43	-5.07	-2.37	-7.58	-2.49	-2.75
β_{21}	-0.41	-0.38	-5.88	-0.429	5.94	-0.41	2.42
β_{31}	0.18	0.16	-5.44	0.19	7.42	0.18	3.74
β_{02}	-2.91	-2.91	0.15	-3.07	5.71	-2.99	3.17
β_{12}	-0.09	-0.08	-3.80	-0.08	-6.16	-0.09	-1.06
β_{22}	0.64	0.63	-1.73	0.67	4.89	0.66	4.08
β_{32}	1.46	1.45	-0.67	1.57	7.06	1.52	4.14
θ	0.49	0.49	0.99	0.51	4.96	0.50	2.17

Unweighted Analyses

Parameter	"True" Parameter	SRS	Relative Bias (%)	Jackknife	Relative Bias (%)	Bootstrap	Relative Bias (%)
β_{01}	6.94	6.37	-8.27	7.08	2.07	7.02	1.16
β_{11}	-2.56	-2.39	-6.83	-2.93	14.30	-2.81	9.69
β_{21}	-0.41	-0.38	-5.89	-0.39	-2.51	-0.39	-2.51
β_{31}	0.17	0.16	-7.65	0.21	14.72	0.19	9.36
β_{02}	-2.91	-2.91	0.15	-3.41	17.29	-2.96	2.17
β_{12}	-0.09	-0.08	-13.51	-0.04	-53.15	0.09	-203.86
β_{22}	0.64	0.63	-1.73	0.73	14.83	0.65	1.51
β_{32}	1.46	1.45	-0.67	1.77	20.68	1.56	6.89
θ	0.49	0.46	-5.07	0.47	-4.61	0.48	-2.69

.

Table A5: Parameter Estimates and Relative Bias

$\theta_1 \sim \beta[70,30]$; $\theta_2 \sim \beta[50,50]$

Weighted Analyses

Parameter	"True" Parameter	SRS	Relative Bias (%)	Jackknife	Relative Bias (%)	Bootstrap	Relative Bias (%)
β_{01}	7.36	7.22	-1.88	7.24	-1.64	7.25	-1.50
β_{11}	-2.73	-2.67	-2.28	-2.87	4.96	-2.87	4.86
β_{21}	-0.34	-0.36	3.95	-0.34	-2.56	-0.34	-1.06
β_{31}	0.21	0.19	-5.67	0.20	-4.59	0.20	-2.76
β_{02}	-3.15	-3.01	-4.53	-3.08	-2.35	-3.09	-1.86
β_{12}	-0.07	-0.07	6.51	-0.08	5.16	-0.07	2.82
β_{22}	0.69	0.67	-2.25	0.72	4.76	0.72	3.23
β_{32}	1.59	1.53	-3.66	1.55	-2.70	1.56	-2.14
θ	0.66	0.65	-0.24	0.69	4.65	0.66	0.11

Unweighted Analyses

Parameter	"True" Parameter	SRS	Relative Bias (%)	Jackknife	Relative Bias (%)	Bootstrap	Relative Bias (%)
β_{01}	7.36	7.62	3.56	7.82	6.18	7.90	7.31
β_{11}	-2.73	-2.75	0.52	-3.04	11.26	-2.99	9.29
β_{21}	-0.34	-0.39	12.09	-0.39	13.04	-0.39	13.00
β_{31}	0.21	0.22	5.37	0.24	16.36	0.24	17.10
β_{02}	-3.15	-2.8	-9.24	-2.87	-8.69	-3.55	12.72
β_{12}	-0.07	-0.08	11.90	-0.08	10.27	-0.08	9.02
β_{22}	0.69	0.62	-9.97	0.62	-9.60	0.64	-8.01
β_{32}	1.59	1.42	-10.75	1.43	-10.23	1.45	-9.23
θ	0.66	0.60	-8.62	0.60	-8.65	0.61	-8.24

Table A6: Parameter Estimates and Relative Bias

$\theta_1 \sim \beta[90,10]$; $\theta_2 \sim \beta[40,60]$

Weighted Analyses

Parameter	"True" Parameter	SRS	Relative Bias (%)	Jackknife	Relative Bias (%)	Bootstrap	Relative Bias (%)
β_{01}	7.56	8.06	6.72	8.05	6.49	8.02	6.11
β_{11}	-2.82	-3.05	8.02	-3.02	7.08	-3.02	7.27
β_{21}	-0.35	-0.36	3.37	-0.35	-2.22	-0.35	-0.64
β_{31}	0.21	0.22	4.38	0.22	4.73	0.22	2.56
β_{02}	-3.23	-3.19	-1.15	-3.17	-1.78	-3.19	-1.31
β_{12}	-0.06	-0.06	1.59	-0.06	0.02	-0.06	0.03
β_{22}	0.68	0.66	-4.18	0.66	-4.32	0.66	-2.96
β_{32}	1.62	1.67	2.98	1.66	2.38	1.65	1.77
θ	0.79	0.79	-0.22	0.82	3.52	0.78	-0.52

Unweighted Analyses

Parameter	"True" Parameter	SRS	Relative Bias (%)	Jackknife	Relative Bias (%)	Bootstrap	Relative Bias (%)
β_{01}	7.56	8.07	6.82	8.07	6.71	8.03	6.25
β_{11}	-2.82	-2.72	-3.51	-2.93	3.74	-2.96	5.04
β_{21}	-0.35	-0.34	-4.78	-0.35	0.39	-0.4	12.45
β_{31}	0.21	0.20	-6.95	0.18	-13.33	0.20	-2.89
β_{02}	-3.23	-2.98	-7.82	-3.16	-2.14	-3.67	13.62
β_{12}	-0.06	-0.06	2.39	-0.06	10.67	-0.07	9.08
β_{22}	0.68	0.62	-8.71	0.65	-3.98	0.67	-1.06
β_{32}	1.62	1.49	-8.35	1.53	-5.53	1.53	-5.81
θ	0.79	0.65	-17.43	0.65	-17.45	0.65	-17.24

Table A7 "True" and Estimated Standard Errors and Design Effects

$\theta_1 \sim \beta[5,5]$; $\theta_2 \sim \beta[4,6]$

Weighted Analyses

Parameter	"True" SE	SRS	Ratio SRS:True	Jackknife	Ratio Jack:True	$DEFF_{jack}$	Bootstrap	Ratio Boot:True	$DEFF_{boot}$
β_{01}	0.03	0.02	0.89	0.04	1.12	1.25	0.03	1.08	1.21
β_{11}	0.01	0.01	0.90	0.01	1.08	1.20	0.01	1.05	1.16
β_{21}	0.01	0.01	0.91	0.01	1.09	1.20	0.01	1.09	1.19
β_{31}	0.01	0.01	0.92	0.01	1.08	1.17	0.01	1.04	1.13
β_{02}	0.01	0.01	0.92	0.01	1.17	1.26	0.01	1.15	1.24
β_{12}	0.003	0.004	0.94	0.004	1.06	1.12	0.003	1.01	1.06
β_{22}	0.004	0.004	0.92	0.004	1.03	1.11	0.004	1.01	1.09
β_{32}	0.005	0.005	0.95	0.006	1.05	1.09	0.006	1.03	1.07
θ	0.0004	.0003	0.81	0.0004	1.01	1.23	0.0004	1.02	1.25

Table A8 "True" and Estimated Standard Errors and Design Effects

$\theta_1 \sim \beta[7,3]$; $\theta_2 \sim \beta[5,5]$

Weighted Analyses

Parameter	"True" SE	SRS	Ratio SRS:True	Jackknife	Ratio Jack:True	$DEFF_{jack}$	Bootstrap	Ratio Boot:True	$DEFF_{boot}$
β_{01}	0.03	0.03	0.93	0.03	1.04	1.11	0.03	1.01	1.08
β_{11}	0.01	0.01	0.95	0.01	1.11	1.16	0.01	1.02	1.07
β_{21}	0.01	0.01	0.92	0.01	1.12	1.21	0.01	1.06	1.14
β_{31}	0.01	0.01	0.92	0.01	1.18	1.27	0.01	1.07	1.15
β_{02}	0.02	0.02	0.91	0.02	1.12	1.22	0.02	1.06	1.15
β_{12}	0.02	0.02	0.93	0.02	1.11	1.19	0.02	1.05	1.12
β_{22}	0.01	0.01	0.94	0.01	1.20	1.27	0.01	1.21	1.29
β_{32}	0.01	0.01	0.91	0.01	1.11	1.22	0.01	1.07	1.17
θ	0.0004	0.0004	0.95	0.0004	1.03	1.08	0.0005	1.21	1.27

Table A9 "True" and Estimated Standard Errors and Design Effects

$\theta_1 \sim \beta[9,1]$; $\theta_2 \sim \beta[4,6]$

Weighted Analyses

Parameter	"True" SE	SRS	Ratio SRS:True	Jackknife	Ratio Jack:True	$DEFF_{jack}$	Bootstrap	Ratio Boot:True	$DEFF_{boot}$
β_{01}	0.03	0.02	0.94	0.03	1.11	1.17	0.03	1.05	1.11
β_{11}	0.01	0.01	0.86	0.01	1.07	1.25	0.01	1.04	1.21
β_{21}	0.01	0.01	0.96	0.01	1.12	1.16	0.01	1.09	1.13
β_{31}	0.01	0.01	0.91	0.01	1.15	1.26	0.01	1.09	1.18
β_{02}	0.02	0.02	0.94	0.02	1.11	1.18	0.02	1.03	1.09
β_{12}	0.02	0.02	0.95	0.02	1.09	1.15	0.02	1.05	1.09
β_{22}	0.01	0.01	0.96	0.01	1.08	1.12	0.01	1.01	1.04
β_{32}	0.02	0.02	0.93	0.02	1.10	1.19	0.02	1.01	1.08
θ	0.0004	0.0003	0.86	0.0003	0.79	0.93	0.0004	1.05	1.22

Table A10 "True" and Estimated Standard Errors and Design Effects

$\theta_1 \sim \beta[50,50]$; $\theta_2 \sim \beta[40,60]$

Weighted Analyses

Parameter	"True" SE	SRS	Ratio SRS:True	Jackknife	Ratio Jack:True	$DEFF_{jack}$	Bootstrap	Ratio Boot:True	$DEFF_{boot}$
β_{01}	0.03	0.03	0.91	0.03	1.07	1.18	0.03	1.04	1.15
β_{11}	0.01	0.01	0.91	0.01	1.03	1.14	0.01	1.03	1.13
β_{21}	0.01	0.01	0.93	0.01	1.08	1.16	0.01	1.07	1.15
β_{31}	0.01	0.01	0.98	0.01	1.04	1.06	0.01	1.02	1.04
β_{02}	0.01	0.01	0.97	0.01	1.08	1.11	0.01	1.03	1.04
β_{12}	0.003	0.003	0.89	0.004	1.15	1.29	0.004	1.08	1.19
β_{22}	0.004	0.004	0.91	0.004	1.08	1.18	0.004	1.05	1.14
β_{32}	0.006	0.005	0.94	0.006	1.04	1.10	0.006	1.05	1.12
θ	0.0004	0.0003	0.86	0.0004	1.04	1.21	0.0004	1.02	1.18

Table A11 "True" and Estimated Standard Errors and Design Effects

$\theta_1 \sim \beta[70,30]$; $\theta_2 \sim \beta[50,50]$

Weighted Analyses

Parameter	"True" SE	SRS	Ratio SRS:True	Jackknife	Ratio Jack:True	$DEFF_{jack}$	Bootstrap	Ratio Boot:True	$DEFF_{boot}$
β_{01}	0.03	0.03	0.97	0.03	1.08	1.11	0.03	1.04	1.07
β_{11}	0.01	0.01	0.94	0.01	1.08	1.16	0.01	1.04	1.11
β_{21}	0.01	0.01	0.87	0.01	1.11	1.28	0.01	1.06	1.22
β_{31}	0.01	0.01	0.97	0.01	1.09	1.17	0.01	1.04	1.08
β_{02}	0.02	0.02	0.96	0.02	1.06	1.10	0.02	1.03	1.07
β_{12}	0.02	0.02	0.94	0.02	1.10	1.18	0.02	1.05	1.12
β_{22}	0.01	0.01	0.94	0.01	1.14	1.21	0.01	1.09	1.17
β_{32}	0.02	0.02	0.98	0.02	1.05	1.08	0.02	1.02	1.04
θ	0.0004	0.0004	0.99	0.0004	0.97	0.97	0.0004	1.07	1.07

Table A12 "True" and Estimated Standard Errors and Design Effects

$\theta_1 \sim \beta[90,10]$; $\theta_2 \sim \beta[40,60]$

Weighted Analyses

Parameter	"True" SE	SRS	Ratio SRS:True	Jackknife	Ratio Jack:True	$DEFF_{jack}$	Bootstrap	Ratio Boot:True	$DEFF_{boot}$
β_{01}	0.03	0.02	0.97	0.03	1.10	1.13	0.03	1.02	1.05
β_{11}	0.01	0.01	0.98	0.01	1.09	1.10	0.01	1.05	1.06
β_{21}	0.01	0.01	0.92	0.01	1.09	1.18	0.01	1.04	1.13
β_{31}	0.01	0.01	0.99	0.01	1.08	1.08	0.01	1.05	1.05
β_{02}	0.02	0.02	0.95	0.02	1.04	1.09	0.02	1.03	1.07
β_{12}	0.02	0.02	0.96	0.02	1.07	1.11	0.02	1.05	1.08
β_{22}	0.02	0.02	0.95	0.02	1.09	1.15	0.02	1.05	1.09
β_{32}	0.01	0.01	0.95	0.01	1.07	1.13	0.01	1.05	1.09
θ	0.0005	0.0004	0.82	0.0005	1.13	1.36	0.0004	0.92	1.11

Table A13: 95% Confidence Interval Coverage

$\theta_1 \sim \beta[5,5]$; $\theta_2 \sim \beta[4,6]$

Weighted Analyses

Parameter	SRS	Jackknife	Bootstrap	$DEFF_{jack}$	$DEFF_{boot}$
β_{01}	89.7	94.4	96.2	90.2	89.9
β_{11}	87.3	92.7	95.2	89.1	90.4
β_{21}	86.5	91.4	94.8	88.7	88.5
β_{31}	85.4	91.7	94.7	87.4	86.6
β_{02}	85.0	90.4	94.0	86.7	86.4
β_{12}	86.4	90.4	90.3	86.5	86.2
β_{22}	91.1	90.7	93.0	87.7	87.4
β_{32}	88.4	91.7	91.7	91.8	91.5
θ	91.5	91.2	90.5	89.9	90.2

Unweighted Analyses

Parameter	SRS	Jackknife	Bootstrap	$DEFF_{jack}$	$DEFF_{boot}$
β_{01}	92.2	82.7	88.7	93.5	92.4
β_{11}	89.7	84.3	86.3	90.7	89.7
β_{21}	91.5	79.7	80.4	91.8	91.7
β_{31}	75.4	71.4	89.8	76.5	77.7
β_{02}	85.6	88.7	85.8	83.2	86.2
β_{12}	76.5	82.4	89.4	72.9	77.8
β_{22}	85.7	88.6	86.1	85.9	86.5
β_{32}	82.7	87.2	85.7	80.7	83.7
θ	84.7	86.5	89.3	86.2	85.9

Table A14: 95% Confidence Interval Coverage

$\theta_1 \sim \beta[7,3]$; $\theta_2 \sim \beta[5,5]$

Weighted Analyses

Parameter	SRS	Jackknife	Bootstrap	$DEFF_{jack}$	$DEFF_{boot}$
β_{01}	85.4	92.4	93.5	87.2	86.5
β_{11}	85.3	90.7	92.5	87.0	86.7
β_{21}	85.5	91.4	93.7	87.5	86.6
β_{31}	84.9	91.7	93.7	85.6	85.1
β_{02}	86.7	90.7	94.8	87.0	86.8
β_{12}	86.3	91.8	93.9	87.4	86.8
β_{22}	88.0	91.9	92.8	86.9	87.1
β_{32}	86.2	90.5	91.8	88.5	88.2
θ	88.7	90.9	92.5	87.5	88.4

Unweighted Analyses

Parameter	SRS	Jackknife	Bootstrap	$DEFF_{jack}$	$DEFF_{boot}$
β_{01}	85.7	85.1	86.5	86.1	85.9
β_{11}	83.2	82.6	84.3	84.3	83.7
β_{21}	81.7	83.5	88.4	83.1	84.6
β_{31}	73.4	81.4	87.2	73.8	73.6
β_{02}	80.6	80.1	82.7	75.1	82.4
β_{12}	75.7	74.7	86.9	70.7	77.7
β_{22}	76.4	77.8	86.7	74.7	78.4
β_{32}	79.5	74.7	80.5	75.1	80.9
θ	81.7	82.0	83.5	81.7	82.0

Table A15: 95% Confidence Interval Coverage

$\theta_1 \sim \beta[9,1]$; $\theta_2 \sim \beta[4,6]$

Weighted Analyses

Parameter	SRS	Jackknife	Bootstrap	$DEFF_{jack}$	$DEFF_{boot}$
β_{01}	84.2	89.3	92.0	85.3	84.9
β_{11}	84.7	88.3	88.5	85.9	85.7
β_{21}	85.1	89.7	90.7	86.0	85.6
β_{31}	86.9	87.2	89.9	87.6	87.4
β_{02}	86.9	90.7	92.9	87.5	87.3
β_{12}	86.7	88.3	90.7	87.6	87.0
β_{22}	85.3	89.5	91.0	85.9	85.4
β_{32}	85.8	88.0	90.1	86.8	86.4
θ	90.4	85.1	94.4	87.1	90.7

Unweighted Analyses

Parameter	SRS	Jackknife	Bootstrap	$DEFF_{jack}$	$DEFF_{boot}$
β_{01}	90.0	91.5	90.5	91.0	90.7
β_{11}	87.5	90.7	90.1	88.6	88.9
β_{21}	88.5	92.5	93.4	89.0	89.2
β_{31}	75.5	82.0	89.9	76.8	78.4
β_{02}	77.6	85.5	92.2	76.2	78.1
β_{12}	65.4	66.8	77.0	63.4	66.5
β_{22}	74.3	85.1	89.4	72.7	76.0
β_{32}	73.7	83.4	88.5	70.0	74.7
θ	79.4	87.5	87.3	80.9	80.4

Table A16: 95% Confidence Interval Coverage

$\theta_1 \sim \beta[50,50]$; $\theta_2 \sim \beta[40,60]$

Weighted Analyses

Parameter	SRS	Jackknife	Bootstrap	$DEFF_{jack}$	$DEFF_{boot}$
β_{01}	87.4	89.7	91.0	88.5	88.4
β_{11}	87.8	89.0	90.3	89.0	88.9
β_{21}	86.9	88.3	90.5	87.7	87.5
β_{31}	86.4	87.5	89.6	86.5	86.3
β_{02}	92.8	86.9	89.7	93.0	91.8
β_{12}	88.4	86.9	90.7	88.8	89.8
β_{22}	89.4	87.6	89.2	89.8	89.1
β_{32}	89.4	86.4	89.5	89.7	90.1
θ	89.7	90.3	90.7	90.9	90.7

Unweighted Analyses

Parameter	SRS	Jackknife	Bootstrap	$DEFF_{jack}$	$DEFF_{boot}$
β_{01}	84.6	89.5	92.4	86.6	85.4
β_{11}	83.5	84.3	88.7	85.7	84.6
β_{21}	84.7	88.5	89.2	86.1	85.1
β_{31}	83.7	84.6	88.4	84.5	84.2
β_{02}	89.4	81.6	88.7	90.5	90.1
β_{12}	79.4	55.9	35.7	80.7	81.5
β_{22}	88.4	83.7	88.1	89.8	90.1
β_{32}	91.4	80.1	86.3	92.0	91.5
θ	87.9	88.4	87.9	88.3	88.9

Table A17: 95% Confidence Interval Coverage

$\theta_1 \sim \beta[70,30]$; $\theta_2 \sim \beta[50,50]$

Weighted Analyses

Parameter	SRS	Jackknife	Bootstrap	$DEFF_{jack}$	$DEFF_{boot}$
β_{01}	88.6	91.4	91.6	89.2	88.8
β_{11}	88.1	92.1	92.4	88.8	88.5
β_{21}	89.1	91.6	92.0	89.6	89.2
β_{31}	87.6	90.5	91.7	88.4	87.9
β_{02}	86.4	91.0	93.8	87.2	86.8
β_{12}	85.5	92.1	92.4	86.4	85.8
β_{22}	89.4	90.4	91.7	90.5	89.3
β_{32}	88.6	90.7	90.5	89.9	89.5
θ	92.1	90.6	95.3	91.5	92.4

Unweighted Analyses

Parameter	SRS	Jackknife	Bootstrap	$DEFF_{jack}$	$DEFF_{boot}$
β_{01}	85.2	86.5	85.6	85.9	85.6
β_{11}	93.9	81.6	83.6	94.5	94.0
β_{21}	81.6	82.4	83.5	83.5	83.8
β_{31}	84.6	83.2	82.6	85.7	85.1
β_{02}	80.5	79.5	86.5	79.6	81.6
β_{12}	77.6	74.6	72.6	73.5	75.6
β_{22}	76.5	75.8	88.6	74.6	77.3
β_{32}	77.6	69.5	75.4	74.9	76.1
θ	80.5	79.2	84.5	78.9	80.8

Table A18: 95% Confidence Interval Coverage

$\theta_1 \sim \beta[90,10]$; $\theta_2 \sim \beta[40,60]$

Weighted Analyses

Parameter	SRS	Jackknife	Bootstrap	$DEFF_{jack}$	$DEFF_{boot}$
β_{01}	86.5	88.5	88.4	87.9	87.1
β_{11}	84.5	87.5	87.9	85.9	85.2
β_{21}	85.4	88.2	92.0	86.6	86.9
β_{31}	87.5	88.3	89.6	88.2	88.0
β_{02}	87.2	89.6	89.4	88.1	87.8
β_{12}	88.9	93.7	95.7	89.4	89.0
β_{22}	85.3	89.6	90.1	86.4	85.9
β_{32}	88.5	90.1	89.9	90.1	89.5
θ	91.2	89.7	93.5	92.2	91.4

Unweighted Analyses

Parameter	SRS	Jackknife	Bootstrap	$DEFF_{jack}$	$DEFF_{boot}$
β_{01}	90.4	88.6	89.5	92.1	91.5
β_{11}	90.6	89.6	88.6	91.8	90.8
β_{21}	88.5	93.2	87.1	88.9	89.5
β_{31}	89.4	82.4	90.6	90.1	91.7
β_{02}	78.4	86.7	84.6	77.5	80.0
β_{12}	74.6	78.9	88.3	73.2	75.5
β_{22}	71.9	79.5	89.6	69.5	70.6
β_{32}	75.9	81.3	89.4	73.8	76.9
θ	75.8	77.6	76.5	75.4	75.0

A19: Summary of IYTS Analyses						
Parameter	Weighted Parameter Estimate	Unweighted Parameter Estimate	Relative Bias (%)	SRS Standard Error*	Bootstrap Standard Error*	Ratio Boot:SRS *
β_{01}	6.81	6.12	10.13	0.94	0.99	1.06
β_{11}	-2.47	-1.59	35.16	0.27	0.31	1.12
β_{21}	-0.33	0.22	167.97	0.18	0.21	1.20
β_{31}	0.15	0.06	64.28	0.07	0.09	1.30
β_{02}	-2.49	-3.44	-38.94	1.71	1.90	1.11
β_{12}	-0.14	-0.20	-38.78	0.33	0.54	1.04
β_{22}	0.51	0.63	-22.09	0.36	0.38	1.04
β_{32}	1.22	2.00	-64.82	0.32	0.34	1.04
θ	0.82	0.62	24.05	0.02	0.04	1.61

* Refers to weighted analysis

APPENDIX B: SAS PROGRAMS

```
/*Jack & Boot*/
/*This Program performs model based, jackknifem, and bootstrap*/
/*estimation*/
options nofmterr mprint spool;
proc printto print = output
                        log =lr.logs.log.log label = "log";
                              run;

%macro weighted;
%let seed = 121;
%do i = 1 %to 1000;
/*first data step simulates observations for strata one*/
data strata1 (drop = w i);
   seed=&seed;  /*seed for covariate generation*/
call streaminit(35131); /*seed for beta number generation*/
strata = 1; /*identifies which stratum observation are in*/
do w=1 to 75;
psu = w;
 beta = rand('BETA',5,5);/*beta distribution for cluster MEAN = .9 and
ICC = .0099*/
do i=1 to 20;  /*NUMBER OF OBSERVATIONS TO SIMULATE*/
    weight=1;         /*stratum 1 weight*/
      x1=rannor(seed);  /*1ST COVARIATE*/
      x2=rannor(seed);  /*SECOND COVARIATE*/
      x3=rannor(seed); /*third covariate*/

logit1=6.81  -2.5*x1 - .3*x2 + .2*x3; logit2= -2.5 + -.2*x1 + .5*x2 +
1.2*x3;    /*logits for 2 mixture components*/
      p=beta*exp(-logit1)/(1+exp(-logit1))+(1-beta)*exp(-
logit2)/(1+exp(-logit2)); /*mixing proportion*/
      if ranuni(seed)>p then y=1; else y=0;   /*CREATE DICHOTOMOUS DV*/
      output;
end;
end;
run;
%let seed = &seed + 1;
/*second data step simulates observations for strata two*/
data strata2 (drop = w i);
   seed=&seed;  /*seed for covariate generation*/
call streaminit(17111); /*seed for beta number generation*/
strata = 2; /*identifies which stratum observation are in*/
do w=1 to 75;
psu = w;
beta = rand('BETA',4,6); /*beta distribution for cluster MEAN = .4 and
ICC=.0099 */
do i=1 to 20;  /*NUMBER OF OBSERVATIONS TO SIMULATE*/
    weight = 0.25;     /*stratum 2 weight*/
    x1=rannor(seed);  /*1ST COVARIATE*/
      x2=rannor(seed);  /*SECOND COVARIATE*/
      x3=rannor(seed); /*third covariate*/

logit1=6.81  -2.5*x1 - .3*x2 + .2*x3; logit2= -2.5 + -.2*x1 + .5*x2 +
1.2*x3;    /*logits for 2 mixture components*/
```

```
      p=beta*exp(-logit1)/(1+exp(-logit1))+(1-beta)*exp(-
logit2)/(1+exp(-logit2)); /*mixing proportion*/
      if ranuni(seed)>p then y=1; else y=0;   /*CREATE DICHOTOMOUS DV*/
      output;
end;
end;


run;
/*this datastep merges the two strata into one 'superpopulation'*/
data dissPSU (drop = seed);
set strata1 strata2;
run;
/*this section does weighted pseudo likelihood estimation on tesT data
by Newton-Raphson */
/*dissPSU is for model based estimation (no jackknife or bootstrapp)*/
proc nlp data=dissPSU cov=2 OUTEST = TEMP noprint;
   max llik;
   parms a1=6.5,b1=-2,b2=-1,b3=.5,a2=-2.5,b4=-1,b5=.5,b6=1,theta=.8;
/* initial values */
   bounds 0 <= theta <= 1;

q=theta*exp(a1 + b1*x1 + b2*x2 + b3*x3)/(1 + exp(a1 + b1*x1 + b2*x2 +
b3*x3))+(1-theta)*exp(a2 + b4*x1 + b5*x2 + b6*x3)/(1 + exp(a2 + b4*x1 +
b5*x2 + b6*x3)); /* logit mixture */
   llik=(weight*y)*log(q) + (weight*(1-y))*log(1-q); /*weighted
likelihood function*/
run;
%if &i = 1 %then %do;           /*this loop creates a dataset at
iteration 1 */
  data weighted_parms_5_5_4_6 (DROP = _TECH_ _NAME_ _RHS_ _iter_);
/*contains parameters, SE, & teration*/
    set temp;
    iter = &i;
      if _type_ = "PARMS" then output;
  else if _type_ =  "STDERR" then output;
run;
  proc append base=lr.weighted_parm_5_5_4_6
data=weighted_parms_5_5_4_6;
  run;
%end;

%else %do;                      /*this loop appends parms, se, & iter
counter for all subsequent iterations*/
  data weighted_parms&i (DROP = _TECH_ _NAME_ _RHS_ _iter_);
    set temp;
    iter = &i;
      if _type_ = "PARMS" then output;
  else if _type_ =  "STDERR" then output;
run;

  proc append base=lr.weighted_parm_5_5_4_6 data=weighted_parms&i;
  run;
%end;
proc datasets nolist; /*deletes temp data to minimize memory usage*/
delete weighted_parms&i; run;
```

```
/*************************************************************************
*********************/
/*this section of code performs the Jackknife repilcates*/
/*************************************************************************
*********************/
%macro jackknife;
/*preform 25 jacknife replicates*/
%do j = 1 %to 25;
* Create a data set with one obs per PSU for stratum 1;
proc freq data=strata1 noprint;
  tables psu / out=psulist1(drop=count percent);
run;
* Take a simple random sample of PSUs from stratum 1;
proc surveyselect data=psuList1 out=psu_jack_Sample1 method=srs n=74
noprint;
run;

* Get all the obs for each sampled PSU in stratum 1;
data strata1sample;
  merge psu_jack_sample1 (in=sample) strata1(in=all);
  by psu;
  if sample and all;
run;
* Create a data set with one obs per PSU for stratum 2;
proc freq data=strata2 noprint;
  tables psu / out=psulist2(drop=count percent);
run;
* Take a simple random sample of PSUs from stratum 2;
proc surveyselect data=psuList2 out=psu_jack_Sample2 method=srs n=74
noprint;
run;

* Get all the obs for each sampled PSU in stratum 2;
data strata2sample;
  merge psu_jack_sample2(in=sample) strata2(in=all);
  by psu;
  if sample and all;
run;

data jack;
set strata1sample strata2sample;
if strata = 1 then weight = 1.0135135;
else if strata = 2 then weight = 0.02533783783;
run;

/*this section does weighted pseudo likelihood estimation on tesT data
by Newton-Raphson */
/*jack is for design based estimation (jackknife)*/
proc nlp data=jack cov=2 OUTEST = TEMP noprint;
   max llik;
   parms a1=6.5,b1=-2,b2=-1,b3=.5,a2=-2.5,b4=-1,b5=.5,b6=1,theta=.8;
/* initial values */
   bounds 0 <= theta <= 1;

q=theta*exp(a1 + b1*x1 + b2*x2 + b3*x3)/(1 + exp(a1 + b1*x1 + b2*x2 +
b3*x3))+(1-theta)*exp(a2 + b4*x1 + b5*x2 + b6*x3)/(1 + exp(a2 + b4*x1 +
b5*x2 + b6*x3)); /* logit mixture */
```

```
   llik=(weight*y)*log(q) + (weight*(1-y))*log(1-q); /*weighted
likelihood function*/
run;


%if &j = 1 %then %do;          /*this loop creates a dataset at
iteration 1 */
  data jackparms_5_5_4_6 (DROP = _TECH_ _NAME_ _RHS_ _iter_);
/*contains parameters, SE, & teration*/
    set temp;
    iter = &j;
    loop = &i;
      if _type_ = "PARMS" then output;


  run;
  proc append base=lr.jackparm_5_5_4_6 data=jackparms_5_5_4_6;
  run;
%end;

%else %do;/*this loop appends parms, se, & iter counter for all
subsequent iterations*/
  data jackparms&j (DROP = _TECH_ _NAME_ _RHS_ _iter_);
    set temp;
    iter = &j;
      loop = &i;
      if _type_ = "PARMS" then output;


  run;

  proc append base=lr.jackparm_5_5_4_6 data=jackparms&j;
  run;
%end;
proc datasets nolist; /*deletes temp data to minimize memory usage*/
delete jackparms&j; run;
%end;

%mend;
%jackknife;

/*************************************************************************
*********************/
/*this section of code performs the Bootstrap repilcates*/
/*************************************************************************
*********************/
%macro Bootstrap;
/*preform 200 Bootstrap replicates*/
%do j = 1 %to 100;

* Create a data set with one obs per PSU for stratum 1;
proc freq data=strata1 noprint;
  tables psu / out=psulist1(drop=count percent);
run;
* Take a simple random sample of PSUs from stratum 1;
proc surveyselect data=psuList1 out=psu_Boot_Sample1 method=urs n=75
noprint outhits;
```

```
run;

* Get all the obs for each sampled PSU in stratum 1;
data strata1sample (drop = numberhits);
  set psu_boot_sample1;
  Drop PSU2;
  do k = 1 to N;
    set strata1 (rename=(PSU=PSU2)) point=k nobs=N;
    if PSU=PSU2 then output;
    end;
run;
* Create a data set with one obs per PSU for stratum 2;
proc freq data=strata2 noprint;
  tables psu / out=psulist2(drop=count percent);
run;
* Take a simple random sample of PSUs from stratum 2;
proc surveyselect data=psuList2 out=psu_Boot_Sample2 method=urs n=75
noprint outhits;
run;

* Get all the obs for each sampled PSU in stratum 2;
data strata2sample (drop = numberhits);
  set psu_boot_sample2;
  Drop PSU2;
  do k = 1 to N;
    set strata2 (rename=(PSU=PSU2)) point=k nobs=N;
    if PSU=PSU2 then output;
    end;
run;

data Boot;
set strata1sample strata2sample;
if strata = 1 then weight = 1;
else if strata = 2 then weight = 0.25;
run;


/*this section does weighted pseudo likelihood estimation on tesT data
by Newton-Raphson */
/*Boot is for design based estimation (Bootstrap)*/
proc nlp data=Boot cov=2 OUTEST = TEMP noprint;
   max llik;
   parms a1=6.5,b1=-2,b2=-1,b3=.5,a2=-2.5,b4=-1,b5=.5,b6=1,theta=.8;
/* initial values */
   bounds 0 <= theta <= 1;

q=theta*exp(a1 + b1*x1 + b2*x2 + b3*x3)/(1 + exp(a1 + b1*x1 + b2*x2 +
b3*x3))+(1-theta)*exp(a2 + b4*x1 + b5*x2 + b6*x3)/(1 + exp(a2 + b4*x1 +
b5*x2 + b6*x3)); /* logit mixture */
   llik=(weight*y)*log(q) + (weight*(1-y))*log(1-q); /*weighted
likelihood function*/
run;


%if &j = 1 %then %do;          /*this loop creates a dataset at
iteration 1 */
```

```
  data Bootparms_5_5_4_6 (DROP = _TECH_ _NAME_ _RHS_ _iter_);
/*contains parameters, SE, & teration*/
    set temp;
    iter = &j;
    loop = &i;
      if _type_ = "PARMS" then output;


  run;
  proc append base=lr.Bootparm_5_5_4_6 data=Bootparms_5_5_4_6;
  run;
%end;

%else %do;/*this loop appends parms, se, & iter counter for all
subsequent iterations*/
  data Bootparms&j (DROP = _TECH_ _NAME_ _RHS_ _iter_);
    set temp;
    iter = &j;
      loop = &i;
      if _type_ = "PARMS" then output;


  run;

  proc append base=lr.Bootparm_5_5_4_6 data=Bootparms&j;
  run;
%end;
proc datasets nolist; /*deletes temp data to minimize memory usage*/
delete Bootparms&j; run;
%end;

%mend;
%Bootstrap;


%end;

%mend;

%weighted;
```

```
/*population generater*/
options nofmterr mprint spool;
proc printto print = output
                        log =lr.logs.log.log label = "log";
                              run;

%macro diss;
%let seed = 175751;
%do i = 1 %to 1000;

/*first data step simulates observations for strata one*/
data strata1 (drop = w i);
   seed=&seed;  /*seed for covariate generation*/
call streaminit(3131); /*seed for beta number generation*/
strata = 1; /*identifies which stratum observation are in*/
do w=1 to 75;
psu = w;
 beta = rand('BETA',7,3);/*beta distribution for cluster MEAN = .7 and
ICC = .0099*/
do i=1 to 20;  /*NUMBER OF OBSERVATIONS TO SIMULATE*/
    weight=1;         /*stratum 1 weight*/
      x1=rannor(seed);  /*1ST COVARIATE*/
      x2=rannor(seed);  /*SECOND COVARIATE*/
      x3=rannor(seed); /*third covariate*/

logit1=6.81  -2.5*x1 - .3*x2 + .2*x3; logit2= -2.5 + -.2*x1 + .5*x2 +
1.2*x3;    /*logits for 2 mixture components*/
      p=beta*exp(-logit1)/(1+exp(-logit1))+(1-beta)*exp(-
logit2)/(1+exp(-logit2)); /*mixing proportion*/
      if ranuni(seed)>p then y=1; else y=0;   /*CREATE DICHOTOMOUS DV*/
      output;
end;
end;
run;
%let seed = &seed + 1;
/*second data step simulates observations for strata two*/
data strata2 (drop = w i);
   seed=&seed;  /*seed for covariate generation*/
call streaminit(1372); /*seed for beta number generation*/
strata = 2; /*identifies which stratum observation are in*/
do w=1 to 75;
psu = w;
beta = rand('BETA',5,5); /*beta distribution for cluster MEAN = .5 and
ICC=.0099 */
do i=1 to 20;  /*NUMBER OF OBSERVATIONS TO SIMULATE*/
    weight = 0.25;     /*stratum 2 weight*/
    x1=rannor(seed);  /*1ST COVARIATE*/
      x2=rannor(seed);  /*SECOND COVARIATE*/
      x3=rannor(seed); /*third covariate*/


logit1=6.81  -2.5*x1 - .3*x2 + .2*x3; logit2= -2.5 + -.2*x1 + .5*x2 +
1.2*x3;    /*logits for 2 mixture components*/
      p=beta*exp(-logit1)/(1+exp(-logit1))+(1-beta)*exp(-
logit2)/(1+exp(-logit2)); /*mixing proportion*/
      if ranuni(seed)>p then y=1; else y=0;   /*CREATE DICHOTOMOUS DV*/
```

```
      output;
end;
end;


run;
/*this datastep merges the two strata into one 'superpopulation'*/
data dissPSU (drop = seed);
set strata1 strata2;
run;
/*this section does weighted pseudo likelihood estimation on tesT data
by Newton-Raphson */
proc nlp data=dissPSU cov=2 OUTEST = TEMP noprint;
   max llik;
   parms a1=6.5,b1=-2,b2=-1,b3=.5,a2=-2.5,b4=-1,b5=.5,b6=1,theta=.8;
/* initial values */
   bounds 0 <= theta <= 1;

q=theta*exp(a1 + b1*x1 + b2*x2 + b3*x3)/(1 + exp(a1 + b1*x1 + b2*x2 +
b3*x3))+(1-theta)*exp(a2 + b4*x1 + b5*x2 + b6*x3)/(1 + exp(a2 + b4*x1 +
b5*x2 + b6*x3)); /* logit mixture */
   llik=(weight*y)*log(q) + (weight*(1-y))*log(1-q); /*weighted
likelihood function*/
run;


%if &i = 1 %then %do;           /*this loop creates a dataset at
iteration 1 */
  data weighted_parms_7_3_5_5 (DROP = _TECH_ _NAME_ _RHS_ _iter_);
/*contains parameters, SE, & teration*/
    set temp;
    iter = &i;
      if _type_ = "PARMS" then output;
    else if _type_ =  "STDERR" then output;

  run;
  proc append base=lr.weighted_parm data=weighted_parms_7_3_5_5;
  run;
%end;

%else %do;                      /*this loop appends parms, se, & iter
counter for all subsequent iterations*/
  data weighted_parms&i (DROP = _TECH_ _NAME_ _RHS_ _iter_);
    set temp;
    iter = &i;
      if _type_ = "PARMS" then output;
    else if _type_ =  "STDERR" then output;

  run;

  proc append base=lr.weighted_parm_7_3_5_5 data=weighted_parms&i;
  run;
%end;
proc datasets nolist; /*deletes temp data to minimize memory usage*/
delete weighted_parms&i; run;
%end;
```

```
%mend;

%diss;
```

REFERENCES

Akaike, H. (1981). Likelihood of a Model and Information Criteria. Journal of Econometrics, 16, 3-14.

Akaike, H. (1974). A New Look at Statistical Model Identification. IEEE Transactions on Automatic Control, 19, 716-723.

Box, G. E., Jenkins, G. M., & Reinsel, G. C. (1994). Time Series Analysis: Forecasting and Control 3rd Edition. Englewood Cliffs, NJ: Prentice Hall.

Brier, S. S. (1980). Analysis of contingency tables under cluster sampling. Biometrika, 67, 591-596.

Cochran, W. G. (1977). Sampling Techniques. New York: John Wiley.

DeSarbo, W. S. & Cron, W. L. (1988). A maximum likelihood methodology for clusterwise linear regression. Journal of Classification, 5, 249-282.

Efron, B. (1979). Bootstrap methods: another look at the jackknife. Annals of Statistics, 7, 1-26.

Efron, B. (1981). Nonparametric estimates of standard error: the jackknife, the bootstrap, and other methods. Biometrika, 68, 589-599.

Efron, B. & Tibshirani, R. J. (1993). An Introduction to the Bootstrap. London: Chapman and Hall.

Everitt, B. S. & Hand, D. J. (1981). Finite Mixture Distributions. London: Chapman and Hall.

Fay, R. E. (1985). A jackknifed chi-squared test for complex samples. Journal of the American Statistical Association, 80, 148-157.

Follman, D. A. & Lambert, D. (1989). Generalizing logistic regression by nonparametric mixing. Journal of the American Statistical Association, 84, 295-300.

Hasselblad, V. (1966). Estimation of parameters for a mixture of normal distributions. Technometrics, 8, 431-444.

Hasselblad, V. (1969). Estimation of finite mixtures of distributions from the exponential family. Journal of the American Statistical Association, 64, 1459-1471.

Hosmer, D. W. & Lemeshow, S. (2000). Applied Logistic Regression: 2nd Ed. New York: John Wiley and Sons

John, S. (1970). On analyzing mixed samples. Journal of the American Statistical Association, 65, 755-760.

Kish, L. (1965). Survey Sampling. New York: John Wiley and Sons.

Kleinbaum, D. G., Kupper, L. L., & Muller, K. E. (1988). Applied Regression Analysis and Other Multivariable Methods. Boston: PWS-Kent.

Korn, E. L. & Graubard, B. I. (1991). Epidemiologic studies utilizing surveys: accounting for the sampling design. Journal of the Royal Statistical Society A., 158, 263-295.

Lohr, S. L. (1999). Sampling: Design and Analysis. New York: Duxbury Press.

Lwin, T. & Martin, P. J. (1989). Probits of mixtures. Biometrics, 45, 721-732.

McCullagh, P. & Nelder, J. A. (1989), Generalized Linear Models. New York: Chapman and Hall.

Mosteller, F. & Tukey, J. W. (1968). Data analysis, including statistics. In G. Lindsay and E. Aronson (Eds.) The Handbook of Social Psychology, (Vol. 2). Reading, MA: Addison-Wesley. 80-203.

Newcomb, S. (1886). A generalized theory of the combination of observations so as to obtain the best result. American Journal of Mathematics,8, 343-366.

Patterson, B. H. (1998). Latent class analysis of sample survey data. Ph.D. dissertation, University of Maryland, Department of Measurement, Statistics, and Evaluation.

Patterson, B. H., Dayton, C. M., & Graubard, B. L. (2002). Latent class analysis of complex survey data: Application to dietary data. Journal of the American Statistical ssociation, 97, 1-9.

Peasron, K. (1894). Contributions to the mathematical theory of evolution. Philosophical Transactions of the Royal Society *A*, 185, 71-110.

Pedhazur, E. J. (1997). Multiple Regression in Behaviorl Research. Fort Worth, TX: Harcourt Brace.

Quandt, R. E. & Ramsey, J. B. (1978). Estimating mixtures of normal distributions and switching regressions. Journal of the American Statistical Association, 73, 730-738.

Quenouille, M. H. (1949). Approximate tests of correlation in time series. Journal of the Royal Statistical Society B, 11, 68-84.

Quenouille, M. H. (1956). Notes on bias in estimation. Biometrika, 43, 353-360.

Ramaswamy, V., Anderson, E., & DeSarbo, W. S., (1994). Clusterwise

Negative Binomial Models For Count Data. Management Science, 40, 405-412.

Roberts, G., Rao, J. N. K., & Kumar, S. (1987). Logistic regression analysis of sample survey data. Biometrika, 74, 1-12.

Rust, K. F., & Rao, J. N. K. (1996). Variance estimation for complex surveys using replication techniques. Statistical Methods in Medical research, 5, 283-310.

SAS Institute Inc. (1999). SAS/OR User's Guide: Mathematical Programming, Version 8. Cary, NC: SAS Institute Inc.

Schwarz, G. (1978). Estimating the dimension of a model. Annals of Statistics, 6, 461-464.

Shao, J. & Tu, D. (1995). The jackknife and bootstrap. New York: Springer.

Skinner, C. J., Holt, D., & Smith, T. M. F. (1989). Analysis of Complex Surveys. New York: John Wiley and Sons.

Thomas, D. R. & Rao, J. N. K. (1987). Small-sample comparisons of level and power for simple goodness-of-fit statistics under cluster sampling. Journal of the American Statistical Association, 82, 630-636.

Titterington, D. M., Smith, A. F. M., & Makov, U. E. (1985). Statistical Analysis of Finite Mixture Distributions. New York: John Wiley and Sons.

Tukey, J. W. (1958). Bias and confidence in not quite large samples. Annals of Mathematical Statistics, 29, 614.

Wedel, M. & DeSarbo, W. S. (1992). A latent class binomial logit methodology for the analysis of paired comparison choice data. Memorandum from Institute of

Economic Research, nr. 467, Faculty of Economics, University of Groningen, Netherlands.

Wedel, M. & DeSarbo, W. S. (2002). A Review of Recent Developments in Latent Class Regression Models. In R. Bagozzi (Ed.). Advanced Methods of Marketing Research. London: Blackwell. 352-388.

Wedel, M., DeSarbo, W. S., Bult, J. R., & Ramaswamy, V. (1991). A latent class Poisson regression model for heterogeneous count data. Memorandum from Institute of Economic Research, nr. 470, Faculty of Economics, University of Groningen, Netherlands.