

ACCOUNTING FOR NON-GAUSSIAN MEASUREMENT ERROR IN COMPLEX SURVEY ESTIMATORS OF DISTRIBUTION FUNCTIONS AND QUANTILES

John L. Eltinge

Texas A&M University

Abstract: In the analysis of complex survey data, it is often important to estimate distribution functions or quantiles associated with a given variable x . Estimation of these values is known to be problematic if the variable x is measured with error. Previous work with this problem generally has used the assumption that the measurement errors, or transformations thereof, have a normal distribution. This paper uses small-error approximations to develop adjusted estimators of distribution functions or quantiles for cases in which measurement errors are nonnormal. These approximations also lead to some relatively simple diagnostics that indicate the extent to which customary sample survey distribution function estimators are sensitive to: (1) varying magnitudes of measurement error; and (2) the approximate shape of the distribution of the true x values. Some of the proposed diagnostics require identification information, e.g., estimates of the measurement error variance, but do not necessarily require direct access to individual-level replicate observations. The proposed methods are applied to data from the U.S. Third National Health and Nutrition Examination Survey (NHANES III).

Key words and phrases: Convolution, misclassification, nonnormal distribution, percentile, prevalence rate, response error, sensitivity analysis, small-error approximation, stratified multistage sample survey, superpopulation model, Third National Health and Nutrition Examination Survey (NHANES III).

1. Introduction

1.1. Distribution functions, prevalence rates and quantiles

In the analysis of complex survey data, one often needs to estimate distribution functions or quantiles associated with specific variables Y . The design-based survey literature has developed a variety of methods for point estimation, variance estimation and confidence interval construction for these parameters, both in a pure finite-population context and in a superpopulation-model context. See, e.g., Woodruff (1952), Rao Kovar and Mantel (1990), Francisco and Fuller (1991), Shao and Wu (1992) and references cited therein.

To review some of the main ideas in this literature, assume that a finite population of size N is generated through a superpopulation with cumulative

distribution function $F_Y(\cdot)$. One then selects a stratified multistage sample, and computes $\hat{F}_Y(x)$, an appropriately weighted sample proportion of units i that have observed Y_i less than or equal to a specified value x . In some cases, a certain prespecified cutoff value x^* is of special substantive interest, and so analysts focus attention on an estimated prevalence rate $\hat{F}_Y(x^*)$, the weighted proportion of sample units i with $Y_i \leq x^*$.

In other cases, analytic interest focuses on the quantiles associated with $F_Y(\cdot)$, defined by

$$x_{F\gamma} = \inf\{x : F_Y(x) \geq \gamma\} \quad \text{for } \gamma \in (0, 1). \quad (1.1)$$

See, e.g., Rao, Kovar and Mantel (1990, p.372). A point estimator $\hat{x}_{F\gamma}$ can be computed by replacing $F_Y(x)$ with $\hat{F}_Y(x)$ in expression (1.1). The quantile estimators $\hat{x}_{F\gamma}$, $\gamma \in (0, 1)$ provide useful descriptive statistics. In addition, there is often strong substantive interest in specific quantiles like the median ($\gamma = 0.5$), quartiles ($\gamma = 0.25, 0.75$) and certain tail quantiles ($\gamma = 0.05$ or 0.95 , say).

Under large-sample conditions, the survey literature has also developed methods for the construction of variance estimators and confidence intervals for the distribution functions and quantiles of interest. For example, for a given point x the estimator $\hat{F}_Y(x)$ generally can be written as a ratio of estimated population totals, and customary variance estimation and confidence interval methods can be applied accordingly. In addition, Woodruff (1952) noted that the inverse relationship between $\hat{x}_{F\gamma}$ and $\hat{F}_Y(x)$, combined with standard confidence bounds for $F_Y(x)$, lead to relatively simple confidence intervals for $x_{F\gamma}$. See, e.g., Francisco and Fuller (1991) for a rigorous review and extension of these ideas. In addition, see Rao, Kovar and Mantel (1990) for some related methods involving the use of auxiliary data.

1.2. An example: prevalence rates and quantiles estimated from NHANES III

The present work was motivated by applications to the U.S. Third National Health and Nutrition Examination Survey (NHANES III). Details of the application will be presented in Section 6, but for the moment we note that NHANES III selected sample persons through a stratified multistage sample design intended to cover the noninstitutionalized U.S. civilian population. Selected sample persons were asked to provide a blood sample and to participate in an interview and a thorough medical examination. This is in contrast with most other national-level sample surveys on health, which generally have not collected blood sample or medical examination variables.

Many of the resulting NHANES III variables are recorded on a continuous scale. A person with a true value Y below a specified cut-point x^* is defined to

have a deficiency disease associated with Y . Consequently, NHANES III analysts have considerable interest in estimation of the associated disease prevalence rates $F_Y(x^*)$. In addition, there is related substantive interest in comparison of quantiles $x_{F\gamma}$ across subpopulations or comparison of $x_{F\gamma}$ values to the quantiles of a reference distribution.

1.3. The effect of measurement error

The abovementioned literature uses the implicit assumption that the observations Y are recorded without error. However, in many cases this assumption is problematic, due to imperfections in the measurement instrument or due to definitional issues. For example, many of the NHANES III variables referred to in Section 1.2 are measured imperfectly.

To develop some notation, consider a sample element i and let Y_i and X_i be the associated “true value” and recorded observation, respectively. Then $X_i - Y_i$ is defined to be the measurement error associated with unit i . If these errors are nontrivial, then it is well known that classical distribution function and quantile estimators can have a correspondingly nontrivial bias.

Consequently, several authors have developed estimators of F_Y and $x_{F\gamma}$ intended to account for the presence of measurement error. For example, Gaffey (1959), building on previous work by Eddington (1913), and Pollard (1953), developed an infinite-series expression for $F_Y(x)$ under the assumption that the measurement errors were normally distributed. This expression involved the distribution function of X evaluated at a specialized set of points in the neighborhood of x ; an infinite number of its derivatives at the same points; and the variances of the measurement errors. Under the assumptions of simple random sampling and known error variances, Gaffey (1959) used his infinite-series result to develop an estimator of $F_Y(x)$ based on the sample distribution function of X , and on second and higher-order differences.

In addition, several application areas have addressed the estimation of distribution functions or prevalence rates in the presence of measurement error. These applications also generally have used the assumption of normal errors, and sometimes have used the assumption of normal true values. See, e.g., quality-control work by Mee Owen and Shyu (1986) and references cited therein.

More recently, Stefanski and Bay (1996) considered cases in which true values arose from either a finite population or from an infinite population with unspecified distribution. They proposed a simulation-extrapolation method to estimate $F_Y(\cdot)$ based on the observed values X collected through a complex sample design. Their development used the assumption that the measurement errors were normal with mean zero and constant variance; and that this variance was either known or estimated from an independent source. Also, several authors

have examined the related problem of estimation of the density of Y based on the observations X . See, e.g., Carroll, Ruppert and Stefanski (1995, Section 12.1) and Stefanski and Bay (1996) for discussion of this literature.

Finally, note that for many applications, the assumption of normal measurement errors can be problematic. To some degree this problem can be addressed by using a monotone transformation $t(\cdot)$ such that the difference $t(X) - t(Y)$ follows an approximate normal distribution. For example, Nusser, Carriquiry, Dodd and Fuller (1996) used a four step estimation method involving preliminary adjustment for nuisance effects; identification of a transformation $t(\cdot)$ such that the population distributions of both $t(Y)$ and $t(X) - t(Y)$ were believed to be approximately normal; use of a normal variance component model to estimate the distribution function of $t(Y)$; and back-transformation to estimate the distribution function of the original Y . However, as noted by Nusser Carriquiry, Dodd, and Fuller (1996, pp.1443-1444), construction of an appropriate transformation is labor-intensive because customary simple transformations (e.g., the Box and Cox (1964) family of transformations) may not suffice. Eckert, Carroll and Wang (1997, Section 3.1) also commented on this problem, and carried out a detailed study of estimation of a transformation $t(\cdot)$ such that the difference $t(X) - t(Y)$ is independent of Y . They emphasized cases in which $t(X) - t(Y)$ has a symmetric distribution, but did not generally require normality of either $t(X) - t(Y)$ or $t(Y)$.

1.4. Adjustment methods and sensitivity analyses to account for non-normal errors and limited information

This paper extends the work reviewed in Section 1.3 by developing methods for the estimation of true-value distribution functions and quantiles in the presence of *non-normal* measurement error. First, Section 2 shows that if measurement errors are relatively small in magnitude, then the superpopulation distribution function of the “noisy” observations X can be approximated by the true-value superpopulation distribution function $F_Y(\cdot)$ and an additional term that depends only on the second derivative of $F_Y(\cdot)$ and on the variance of the measurement errors. This in turn leads to an approximation of $F_Y(\cdot)$ as a function of the observed-value distribution function and its first and second differences. Section 2 also examines the extent to which the approximations are improved under the more restrictive assumption of symmetric measurement errors. Section 3 develops related approximations for quantiles. Special emphasis is placed on a shift function that characterizes the relationship between the quantiles of the X and Y distributions, respectively.

Second, Section 4 uses the results of Sections 2 and 3 to develop adjusted point estimators, variance estimators and confidence intervals for $F_Y(\cdot)$ and $x_{F\gamma}$

based on data from a complex sample survey. These adjustment methods use an estimator of the measurement error variance that may arise either from replicate observations collected for a subsample of sample units, or from an independent source.

Third, the approximations in Sections 2 and 3 also lead directly to simple sensitivity analyses for the potential effect of measurement error on standard estimators of distribution functions or quantiles. Section 5 discusses this idea and Section 6 applies the resulting sensitivity-analysis methods to hemoglobin data from the U.S. Third National Health and Nutrition Examination Survey. Section 7 reviews the main ideas of the paper and notes some extensions to related problems. The appendix presents proofs of the main results of Sections 2 and 3.

2. The Effect of Measurement Error on Superpopulation Distribution Functions

2.1. Superpopulation conditions

Our adjustment methods and diagnostics are based on an approximate relationship among superpopulation distribution functions for the true values Y , and for terms U associated with measurement error. The following conditions will be used.

- (C.1) The random vector (Y, U) is generated by a superpopulation model ξ . Under this model, Y and U have marginal distribution functions $F_Y(\cdot)$ and $G_U(\cdot)$, respectively. In addition, Y and U are independent, and U has an expectation equal to zero.
- (C.2) The distribution function $F_Y(\cdot)$ is absolutely continuous, is three times differentiable, and the first three derivatives of $F_Y(\cdot)$ are uniformly bounded. The first three derivatives associated with $F_Y(\cdot)$ are denoted $F_Y^{(1)}(\cdot)$, $F_Y^{(2)}(\cdot)$ and $F_Y^{(3)}(\cdot)$, respectively.
- (C.3) The third absolute moment associated with $G_U(\cdot)$ is finite.

The assumption in (C.1) of mean-zero measurement error independent of Y is frequently used in measurement error work. The differentiability condition (C.2) is not minimal, but allows a relatively simple local approximation for $F_Y(\cdot)$. Similarly, the bounding conditions in (C.2) and (C.3) are not minimal, but lead to relatively simple approximations for the distribution of X .

2.2. Two small-error approximations

Under condition (C.1), define the observed value

$$X_\delta = Y + \delta U \tag{2.1}$$

where δ is a positive scale factor, and let $H_{X_\delta}(\cdot)$ be the superpopulation cumulative distribution function of X_δ . In addition, recall the standard result (e.g., Woodroffe (1975), p.201) that since $F_Y(\cdot)$ is absolutely continuous, model (2.1) implies that $H_{X_\delta}(\cdot)$ is also absolutely continuous, even though continuity of $G(\cdot)$ is not assumed.

The work below will use “small error” approximations in which δ is assumed to converge to zero. Small-error approaches are frequently used in measurement-error analyses. In keeping with comments by Stefanski and Bay (1996, p.408) and others, it should be emphasized that methods developed from small-error approximations often are useful even when realized measurement errors are moderate; this is somewhat analogous to the applicability of customary large-sample approximations to cases involving moderate sample sizes. However, small-error approximations will not necessarily apply to cases (e.g, Nusser, Carriquiry, Dodd and Fuller (1996), Table 1) in which the variance of δU is of the same or larger magnitude than the variance of Y . In the present superpopulation context, small-error approximations lead to two useful results.

First, we can approximate the superpopulation distribution function of X_δ in a way that depends only on the true-value distribution $F_Y(\cdot)$, the second derivative of $F_Y(\cdot)$ and the superpopulation variance $\delta^2\sigma_U^2$ of the measurement errors δU .

Result 2.1. *Assume conditions (C.1) through (C.3) and define X_δ by expression (2.1). Then as δ converges to zero,*

$$H_{X_\delta}(x) = F_Y(x) + 2^{-1}\{F_Y^{(2)}(x)\}\delta^2\sigma_U^2 + O(\delta^3) \quad (2.2)$$

where σ_U^2 is the variance associated with $G_U(\cdot)$.

The Appendix presents a direct proof of Result 2.1. Note especially that the conditions for Result 2.1 do not require continuity of the measurement errors δU . Several related results have been developed in previous work with continuous measurement errors. For example, expression (2.2) is similar to the first two terms of an infinite-series expansion of a distribution-function convolution in Eddington (1913). See, e.g., Gaffey (1959) for further discussion of such expansions under the assumption of normal measurement errors. Also, under the assumption of normal measurement errors, Stefanski and Bay (1996, expression (9)) gave a small-error approximation for the expectation of a weighted sample distribution function of X_δ . Their approximation was similar to the present expression (2.2) and depended explicitly on $\delta^2\sigma_U^2$, $F_Y(\cdot)$, $F_Y^{(2)}(\cdot)$ and $F_Y^{(4)}(\cdot)$; given a bounded sixth derivative $F_Y^{(6)}(\cdot)$, their approximation error was $O(\delta^6)$. Also, under the assumptions of continuous measurement error, Chesher (1991, expression (2.4)) developed a small-error approximation for the density of X_δ in a multivariate

setting. Chesher (1991, p.456) then used this density result to establish a form of expression (2.2) involving an order $o(\delta^2)$ remainder term.

Second, we can extend Result 2.1 to approximate $F_Y(\cdot)$ as a function of $H_{X\delta}(\cdot)$ and $\delta^2\sigma_U^2$. Specifically, consider the second derivative $F_Y^{(2)}(\cdot)$ in (2.2). For work with distribution functions of Y , some previous work has implicitly approximated the first derivative of F_Y through scaled first differences $(2d)^{-1}\{F_Y(x+d) - F_Y(x-d)\}$; and similar ideas have been used in quantile estimation. See, e.g., Kovar, Rao and Wu (1988), Rao, Kovar and Mantel (1990) and Francisco and Fuller (1991). Result 2.2 uses a similar idea involving second differences of $H_{X\delta}(\cdot)$ to approximate $F_Y^{(2)}(\cdot)$. Section 4 will use this result to construct an estimator of $F_Y(\cdot)$ that adjusts for measurement error.

Result 2.2. *Assume the conditions of Result 2.1, and let c be a fixed positive constant. Then as δ converges to zero,*

$$c^{-2}\{H_{X\delta}(x+c\delta\sigma_U) + H_{X\delta}(x-c\delta\sigma_U) - 2H_{X\delta}(x)\} = F_Y^{(2)}(x)\delta^2\sigma_U^2 + O(\delta^3). \quad (2.3)$$

Consequently,

$$F_Y(x) = H_{X\delta}(x) - 2^{-1}c^{-2}\{H_{X\delta}(x+c\delta\sigma_U) + H_{X\delta}(x-c\delta\sigma_U) - 2H_{X\delta}(x)\} + O(\delta^3). \quad (2.4)$$

Note especially that in Result 2.2, the differencing-distance term $c\delta\sigma_U$ is a constant multiple of our measurement-error standard deviation $\delta\sigma_U$. Thus, in an informal sense, the differencing operation (and thus the approximation of $F_Y^{(2)}(x)$) is carried out in a neighborhood of x , with the width of the neighborhood determined by the magnitude of the measurement errors. Sections 4 and 5 will use Result 2.2 to develop adjusted estimators of $F_Y(\cdot)$ and some related diagnostics. Also, the end of Section 4.2 discusses the choice of an appropriate value for the constant c in applications.

2.3. Related comments

Results 2.1 and 2.2 lead to several related comments. First, note that Result 2.2 approximated $F_Y^{(2)}(x)$ through a relatively simple second difference of $H_{X\delta}(x)$ evaluated at three points. Alternative methods for approximation and estimation of $F_Y^{(2)}(\cdot)$, e.g., local quadratic regression, are also possible, but will not be considered further here.

Second, note that Results 2.1 and 2.2 were stated as superpopulation model results. Analogous, but somewhat more cumbersome, results can be obtained asymptotically for a sequence of finite populations. However, in the interest of space the mathematical development of Sections 2 and 3 is focused on superpopulation results. Similarly, Sections 4 through 6 will restrict attention to estimation and inference for superpopulation parameters.

Finally, note that in some cases (e.g., Eckert, Carroll and wang (1997)) an analyst may believe that the distribution of measurement errors is symmetric but nonnormal. For a symmetric error distribution, we can reduce the magnitude of the approximation errors in Results 2.1 and 2.2. Result 2.3 gives a formal statement of this idea.

Result 2.3. *Assume condition (C.1) and let c be a fixed positive constant. In addition, assume that $F_Y(\cdot)$ is four times differentiable, and that the first four derivatives of $F_Y(\cdot)$ are uniformly bounded. Also, assume that the fourth moment associated with $G_U(\cdot)$ is finite, and that for any $t \in (-\infty, \infty)$,*

$$G_U(t) = 1 - G_U(-t) \quad (2.5)$$

i.e., U is distributed symmetrically around zero. Then as δ converges to zero,

$$H_{X\delta}(x) = F_Y(x) + 2^{-1}\{F_Y^{(2)}(x)\}\delta^2\sigma_U^2 + O(\delta^4) \quad (2.6)$$

and

$$F_Y(x) = H_{X\delta}(x) - 2^{-1}c^{-2}\{H_{X\delta}(x + c\delta\sigma_U) + H_{X\delta}(x - c\delta\sigma_U) - 2H_{X\delta}(x)\} + O(\delta^4). \quad (2.7)$$

Results 2.1 through 2.3, combined with variants on expression (9) of Stefanski and Bay (1996), give a hierarchy of approximations. Specifically, the mild conditions (C.1) through (C.3) lead to an approximation with error of order $O(\delta^3)$; the additional assumption of a symmetric measurement error distribution reduces the approximation error to order $O(\delta^4)$; and the more restrictive assumption of normal measurement error further reduces the approximation error to $O(\delta^6)$.

2.4. Superpopulation models with clustering

Note that condition (C.1) includes the assumption that our finite population was generated through independent and identically distributed realizations of a superpopulation model. This assumption is used fairly commonly; see, e.g., Francisco and Fuller (1991) on distribution functions and quantiles; and Binder (1983) on parameters associated with superpopulation-level estimating equations.

However, in some cases substantive features of a population suggest some correlation among true values Y within groups of units at the superpopulation level. For these cases, the assumption of independent and identically distributed superpopulation realizations should be replaced with a more refined superpopulation model intended to account explicitly for superpopulation-level correlation. Under such models, inference regarding superpopulation parameters becomes more complex. See, e.g., Fuller (1975, Appendix A), Korn and Graubard (1998), and references cited therein. For the present work, generalization of condition

(C.1) to account for correlation within groups of elements would lead to fairly straightforward extensions of the parametric relationships developed under (C.1) in Sections 2.2, 2.3 and 3 below. However, confidence bounds for superpopulation parameters considered in Section 4 would require substantial modification; details of these modifications will be considered elsewhere.

3. The Effect of Measurement Error on Superpopulation Quantiles

The ideas of Section 2 extend directly to superpopulation quantiles. Subsection 3.1 outlines some relevant notation and a technical condition. Subsection 3.2 presents two approximations for relationships between the quantiles of the $F_Y(\cdot)$ and $H_{X_\delta}(\cdot)$ distributions. Subsection 3.3 ties these differences to the idea of a “shift function” used in the general distribution-function literature. Subsection 3.4 examines links with previously developed normal-distribution quantile adjustments.

3.1. Notation and an additional condition

As noted in section 2.2, conditions (C.1) and (C.2) and model (2.1) ensure that both $F_Y(\cdot)$ and $H_{X_\delta}(\cdot)$ are absolutely continuous. Thus, for any $\gamma \in (0, 1)$, the inverses $F_Y^{-1}(\cdot)$ and $H_{X_\delta}^{-1}(\cdot)$ are well-defined, and the quantiles $x_{F_\gamma} \stackrel{\text{def}}{=} \inf\{x : F_Y(x) \geq \gamma\} = F_Y^{-1}(\gamma)$ and $x_{H_{\delta\gamma}} \stackrel{\text{def}}{=} \inf\{x : H_{X_\delta}(x) \geq \gamma\} = H_{X_\delta}^{-1}(\gamma)$ are uniquely defined. In developing approximations for these quantiles, we will use conditions (C.1) through (C.3) as well as the following additional condition.

(C.4) For any $\gamma \in (0, 1)$, $F_Y^{(1)}(x_{F_\gamma})$ is strictly positive.

3.2. Two small-error approximations for quantiles

Result 3.1 presents a small-error approximation for the quantile $x_{H_{\delta\gamma}}$ as a function of x_{F_γ} and the first and second derivatives of $F_Y(\cdot)$ evaluated at x_{F_γ} .

Result 3.1. *Assume conditions (C.1) through (C.4), define X_δ by expression (2.1), and let $H_{X_\delta}(\cdot)$ be the cumulative distribution function of X_δ . Then for any $\gamma \in (0, 1)$, as δ converges to zero,*

$$x_{H_{\delta\gamma}} = x_{F_\gamma} - 2^{-1}\{F_Y^{(1)}(x_{F_\gamma})\}^{-1}\{F_Y^{(2)}(x_{F_\gamma})\}\delta^2\sigma_U^2 + O(\delta^3). \quad (3.1)$$

The proof of Result 3.1 in the appendix uses some of the same arguments as the usual Bahadur (1966) approximation for sample quantiles. (See Francisco and Fuller (1991, Section 3) for detailed discussion of the Bahadur approximation in the context of complex sample surveys.) However, note that $x_{H_{\delta\gamma}}$ and x_{F_γ} are both superpopulation quantiles rather than sample quantiles. Also, note that

expression (3.1) is a variant on general results for the effect of small measurement error on the bias of some nonlinear estimators, e.g., Carroll and Stefanski (1990). Less formally, note that for a given value of $\delta^2\sigma_U^2$, the adjustment term in (3.1) is largest for x values with the largest values of $F_Y^{(2)}(x)/F_Y^{(1)}(x)$, i.e., at the points x for which the distribution function of Y has the largest amount of curvature relative to its slope.

Result 3.1 extends directly to functions of several quantiles. For example, consider the difference $x_{F\gamma_2} - x_{F\gamma_1}$ of quantiles evaluated at two distinct points $\gamma_1, \gamma_2 \in (0, 1)$; a common example is the interquartile range, with $\gamma_2 = 0.75$ and $\gamma_1 = 0.25$. Then Result 3.1 implies that

$$\begin{aligned} & (x_{H\delta\gamma_2} - x_{H\delta\gamma_1}) - (x_{F\gamma_2} - x_{F\gamma_1}) \\ &= -2^{-1}[\{F_Y^{(1)}(x_{F\gamma_2})\}^{-1}F_Y^{(2)}(x_{F\gamma_2}) - \{F_Y^{(1)}(x_{F\gamma_1})\}^{-1}F_Y^{(2)}(x_{F\gamma_1})]\delta^2\sigma_U^2 + O(\delta^3). \end{aligned} \quad (3.2)$$

Thus, depending on the numerical values of the derivatives of $F_Y(\cdot)$, the leading term of expression (3.2) may be, on an absolute scale, either larger or smaller than the individual differences $x_{H\delta\gamma_i} - x_{F\gamma_i}$, or may be negligible, i.e., $O(\delta^3)$. In particular, suppose that $F_Y^{(2)}(x_{F\gamma_1}) > 0$ and $F_Y^{(2)}(x_{F\gamma_2}) < 0$, as would occur for the interquartile range of a unimodal symmetric distribution. Then the leading (i.e., order $O(\delta^2)$) term of the difference (3.2) is the sum of the absolute values of the positive leading term of $(x_{H\delta\gamma_2} - x_{F\gamma_2})$ and the negative leading term of $(x_{H\delta\gamma_1} - x_{F\gamma_1})$.

Finally, Result 3.2 gives a related small-error approximation for $x_{F\gamma}$ as a function of $x_{H\delta\gamma}$ and first and second-order differences of $H_{X\delta}(\cdot)$ in a neighborhood of $x_{H\delta\gamma}$. Sections 4 and 5 will use this result to construct diagnostics for $x_{F\gamma}$.

Result 3.2. *Assume the conditions of Result 3.1 and let c be a fixed positive constant. Then as δ converges to zero,*

$$\begin{aligned} x_{F\gamma} &= x_{H\delta\gamma} + [c^{-1}\{H_{X\delta}(x_{H\delta\gamma} + c\delta\sigma_U) - H_{X\delta}(x_{H\delta\gamma} - c\delta\sigma_U)\}]^{-1} \\ &\quad \times \{H_{X\delta}(x_{H\delta\gamma} + c\delta\sigma_U) + H_{X\delta}(x_{H\delta\gamma} - c\delta\sigma_U) - 2H_{X\delta}(x_{H\delta\gamma})\}\delta\sigma_U + O(\delta^3). \end{aligned} \quad (3.3)$$

3.3. Related shift functions

In an informal sense, Results 2.1 and 3.1 indicated that measurement error causes the distribution of the observed X_δ to be shifted relative to the distribution of the true values Y ; cf. Chesher (1991) and references cited therein. To formalize this idea, note that the general distribution-function literature (e.g., Doksum and Sievers (1976), p.421) has developed the idea of a *shift function* $\Delta(x)$ that

describes the relationship between the quantiles associated with two distribution functions. For the present measurement-error problem, we define

$$\Delta(x) = H_{X\delta}^{-1}\{F_Y(x)\} - x \tag{3.4}$$

so that $H_{X\delta}\{x + \Delta(x)\} = F_Y(x)$. Thus, for any $\gamma \in (0, 1)$, $H_{X\delta}\{x_{F\gamma} + \Delta(x_{F\gamma})\} = F_Y(x_{F\gamma}) = \gamma$, i.e., if $x_{F\gamma}$ is the γ th quantile for $F_Y(\cdot)$, then $x_{F\gamma} + \Delta(x_{F\gamma})$ is the γ th quantile for $H_{X\delta}(\cdot)$.

Then Result 3.1 implies that the shift function (3.4) evaluated at $x = x_{F\gamma}$ is approximately equal to $-2^{-1}\{F_Y^{(1)}(x_{F\gamma})\}^{-1}\{F_Y^{(2)}(x_{F\gamma})\}\delta^2\sigma_U^2$. Similarly, Result 3.2 implies that this same shift function can be approximated by

$$-c^{-1}\{H_{X\delta}(x_{H\delta\gamma} + c\delta\sigma_U) - H_{X\delta}(x_{H\delta\gamma} - c\delta\sigma_U)\}^{-1} \times \{H_{X\delta}(x_{H\delta\gamma} + c\delta\sigma_U) + H_{X\delta}(x_{H\delta\gamma} - c\delta\sigma_U) - 2H_{X\delta}(x_{H\delta\gamma})\}\delta\sigma_U. \tag{3.5}$$

Graphical displays of estimates of these shift functions provide useful diagnostics, and will be illustrated in the NHANES III application in Section 6.

3.4. Links with normal-distribution quantile adjustments

Finally, consider the relationship between the small-error approximation (3.1) and previously developed normal-distribution results. Assume condition (C.1) and assume that $F_Y(\cdot)$ and $G_U(\cdot)$ are distribution functions for normal random variables with, respectively, means μ_Y and 0 and variances σ_Y^2 and σ_U^2 , respectively. Then standard arguments lead to the *exact* result

$$x_{H\delta\gamma} - \mu_Y = (x_{F\gamma} - \mu_Y)(\sigma_{\delta X}/\sigma_Y), \tag{3.6}$$

where $\sigma_{\delta X}^2 = \sigma_Y^2 + \sigma_{\delta X}^2$. See, e.g., Fuller ((1995), p. 124).

Under the same conditions, $\{F_Y^{(1)}(x_{F\gamma})\}^{-1}F_Y^{(2)}(x_{F\gamma}) = -(\sigma_Y)^{-2}(x_{F\gamma} - \mu_Y)$, and so expression (3.1) becomes equivalent to,

$$x_{H\delta\gamma} - \mu_Y = (x_{F\gamma} - \mu_Y)\{1 + \delta^2\sigma_U^2/(2\sigma_Y^2)\} + O(\delta^3). \tag{3.7}$$

Moreover, Taylor expansion arguments show that as δ converges to zero, $\sigma_X/\sigma_Y = 1 + \delta^2\sigma_U^2/(2\sigma_Y^2) + O(\delta^4)$ and so the leading term of expression (3.7) is approximately equal to expression (3.6).

Thus, expression (3.1) can be viewed as an extension of the normal-distribution result (3.6) to cases involving non-normal distributions and small errors. In addition, the preceding arguments suggest that for non-normal F_Y and small measurement errors, the adequacy of the customary adjustment (3.6) will depend on whether $\{F_Y^{(1)}(x_{F\gamma})\}^{-1}F_Y^{(2)}(x_{F\gamma})$ is well approximated by the moment ratio $-\sigma_Y^{-2}(x_{F\gamma} - \mu_Y)$. For moderately nonnormal distributions (e.g., some

t or mixed-normal cases), this adequacy may be problematic primarily in the tails; for markedly nonnormal distributions, this adequacy may be problematic in both the tails and the center of the distribution.

4. Adjustment of Complex Survey Estimators to Account for Measurement Error

The superpopulation results of Sections 2 and 3 lead directly to adjusted estimators and diagnostics for $F_Y(\cdot)$ and $x_{F\gamma}$. Section 4.1 outlines some preliminary estimators and some general large-sample conditions. Section 4.2 develops an adjusted distribution-function estimator based on an estimator of $\delta^2\sigma_U^2$ obtained from replicate measurements of sample units or alternative sources. Section 4.3 outlines similar estimators for the quantiles $x_{F\gamma}$ and the related shift functions $\Delta(x)$.

4.1. Preliminary estimators and large-sample conditions

The notation used here will build on previous work by Francisco and Fuller (1991). First, let ξ be the basic superpopulation model described by conditions (C.1) through (C.4). Second, consider a sequence of populations, samples and positive scale terms δ_ν indexed by the positive integers ν . For a given ν , we have a population of size N_ν ; each unit in the population is an independent and identically distributed realization of the random vectors $(Y, \delta_\nu U)$. The superpopulation distribution of $Y + \delta_\nu U$ is $H_{\nu X\delta}$, and principal interest involves the superpopulation distribution function F_Y . For convenience, we will use the notation $\sigma_{\nu U\delta}^2 = \delta_\nu^2 \sigma_U^2$. In addition, a complex design leads to a sample of size n_ν . Standard weighted estimation methods lead to the point estimators $\hat{H}_{\nu X\delta}(x)$.

Third, for this section we also assume that we have available an estimator $\hat{\sigma}_{\nu U\delta}^2$ of $\sigma_{\nu U\delta}^2$. In practical applications, this estimator may arise from: (a) replicate measurements of sample units (sometimes known as internal reliability data); (b) comparison of sample measurements to true values (internal validation data); or (c) variance estimates from sources that are independent of the original sample observations (external reliability or validation data). See, e.g., Carroll and Stefanski (1990) for a detailed discussion of these possible data sources. Also, under simple random sampling Zhong (1997) used empirical likelihood methods to combine validation data with error-prone measurements to produce estimators of the distribution function $F_Y(\cdot)$.

In the interest of space, details of the specific cases (a)-(c) are omitted here. Instead, we use the following general framework. For any given k -dimensional real vector $x_{(k)} = (x_1, \dots, x_k)$, let $H_{\nu X\delta}(x_{(k)}) = \{H_{\nu X\delta}(x_1), \dots, H_{\nu X\delta}(x_k)\}$. In addition, for each ν , let $\{\hat{H}_{\nu X\delta}(x_{(k)})', \hat{\sigma}_{\nu U\delta}^2\}'$ be a point estimator for the $(k+1)$ -dimensional

vector $\{H_{\nu X\delta}(x(k))', \sigma_{\nu U\delta}^2\}'$ and let $\hat{V}_{\nu H\sigma}(x(k)) = \hat{V}[\{\hat{H}_{\nu X\delta}(x(k))', \hat{\sigma}_{\nu U\delta}^2\}]$ be an estimator of the $(k+1) \times (k+1)$ covariance matrix of the approximate distribution of $\{\hat{H}_{\nu X\delta}(x(k))', \hat{\sigma}_{\nu U\delta}^2\}'$. We will use the following large-sample condition.

(C.5) As ν increases, N_ν and n_ν increase without bound and the positive terms δ_ν decline to zero. Also, the sequence of random matrices $\{[(\hat{H}_{\nu X\delta}(x(k))', \hat{\sigma}_{\nu U\delta}^2)', \hat{V}_{\nu H\sigma}(x(k))], \nu = 1, 2, \dots\}$ is such that $\{\hat{V}_{\nu H\sigma}(x(k))\}^{-1/2}[\{\hat{H}_{\nu X\delta}(x(k))', \hat{\sigma}_{\nu U\delta}^2\}' - \{H_{\nu X\delta}(x(k))', \sigma_{\nu U\delta}^2\}'] \xrightarrow{d} N(0, I)$, the $(k+1)$ -dimensional normal distribution with mean zero and identity covariance matrix.

Because $\{\hat{H}_{\nu X\delta}(x(k))', \hat{\sigma}_{\nu U\delta}^2\}$ is approximately a weighted mean vector, condition (C.5) is relatively mild. A formal development of sufficient conditions for (C.5) requires a fairly long but routine repetition of standard arguments from, e.g., Krewski and Rao (1981), Francisco and Fuller (1991, especially Theorem 1), Shao (1996) and references cited therein. Consequently, detailed developments of these sufficient conditions are omitted here, and we will make direct use of condition (C.5) itself. Also, to simplify notation the index ν will be omitted in the remainder of this development.

4.2. Distribution functions

Direct substitution of $\hat{H}_{X\delta}(\cdot)$ and $\hat{\sigma}_{U\delta}^2$ into expression (2.4) leads directly to an adjusted point estimator for $F_Y(x)$, defined as

$$\hat{F}_Y(x) = \min[1, \hat{H}_{X\delta}(x) - 2^{-1}c^{-2}\{\hat{H}_{X\delta}(x + c\hat{\sigma}_{U\delta}) + \hat{H}_{X\delta}(x - c\hat{\sigma}_{U\delta}) - 2\hat{H}_{X\delta}(x)\}]. \tag{4.1}$$

Due to the second differences in $\hat{H}_{X\delta}(\cdot)$, the estimator $\hat{F}_Y(x)$ will not necessarily be monotone nondecreasing in x for finite samples. To impose this monotonicity constraint, let $x_1 < x_2 < \dots$ be the distinct values of X observed in a given sample, and define the modified estimator $\tilde{F}_Y(\cdot)$ by

$$\begin{aligned} \tilde{F}_Y(x_1) &= \hat{H}_{X\delta}(x_1); & \tilde{F}_Y(x_i) &= \max\{\hat{F}_Y(x_i), \tilde{F}_Y(x_{i-1})\}, & i \geq 2; \\ & & \text{and } \tilde{F}_Y(x) &= \tilde{F}_Y(x_{i-1}), & x \in (x_{i-1}, x_i). \end{aligned} \tag{4.2}$$

In addition, routine linearization arguments lead to the estimator of the variance of the approximate distribution of $\tilde{F}_Y(x)$, $\hat{V}\{\tilde{F}_Y(x)\} = (d^*)'\hat{V}_{\nu H\sigma}(x_{(3)})(d^*)$, where $d^* = \{d', -2^{-1}\hat{F}^{(2)}(x)\}'$, $d' = (0, 1, 0) - 2^{-1}c^{-2}(1, -2, 1)$, $\hat{F}_Y^{(2)}(x) = (c\hat{\sigma}_{U\delta})^{-2}\{\hat{H}_{X\delta}(x + c\hat{\sigma}_{U\delta}) + \hat{H}_{X\delta}(x - c\hat{\sigma}_{U\delta}) - 2\hat{H}_{X\delta}(x)\}$ and $\hat{V}_{\nu H\sigma}(x_{(3)})$ is the covariance matrix defined by condition (C.5) with $k = 3$ and $x_{(3)} = (x - c\hat{\sigma}_{U\delta}, x, x + c\hat{\sigma}_{U\delta})$. Condition (C.5) and routine approximation arguments then indicate that an approximate $(1 - \alpha)100\%$ confidence interval for $F_Y(x)$ is

$$\{\tilde{F}_{YL}(x), \tilde{F}_{YU}(x)\} = \tilde{F}_Y(x) \pm z_{1-\alpha/2}[\hat{V}\{\tilde{F}_Y(x)\}]^{1/2} \tag{4.3}$$

where $z_{1-\alpha/2}$ is the customary $1 - \alpha/2$ upper quantile of the standard normal distribution.

This estimation work leads to four general comments. First, the point estimator (4.1) is formally similar to the leading terms of a deconvolution estimator proposed by Gaffey (1959, p.199) under the assumption of simple random sampling and normal measurement errors. Second, recall that condition (C.1) required that the true observations Y and the error terms U be independent at the superpopulation level. If this is problematic for a given dataset, one may consider a transformation $t(\cdot)$ such that $t(Y)$ and $t(X) - t(Y)$ are independent at the superpopulation level, in keeping with the discussion of Eckert Carroll and Wang (1997) in Section 1. An estimator of the distribution function of Y then follows from application of (4.2) to the transformed data, followed by back-transformation into the original Y scale. Third, in keeping with case (c) discussed in Section 4.1, consider an estimator $\hat{\sigma}_{U\delta}$ that arises from a source that is independent of $\hat{H}_{X\delta}(x)$ (e.g., through laboratory calibration work separate from the main survey dataset). Then the covariance matrix $\hat{V}_{\nu H\sigma}(x)$ has a block-diagonal form with a null covariance between $\hat{H}_{X\delta}(x)$ and $\hat{\sigma}_{U\delta}$, and computation of $\hat{V}_{\nu H\sigma}(x)$ simplifies accordingly.

Fourth, note that Results 2.2, 2.3 and 3.2 and the estimator (4.1) were developed for an arbitrary fixed positive constant c . For finite sample sizes and nontrivial measurement errors, the choice of an appropriate c will depend on several factors. For example, for a given $F_Y(\cdot)$, x and $\delta\sigma_U$, one would prefer values of c that are small enough to provide a satisfactory approximation (2.3); i.e., such that $F_Y(y)$ and $H_{X\delta}(y)$ are approximately quadratic over the interval $y \in [x - c\delta\sigma_U, x + c\delta\sigma_U]$. On the other hand, for a given $F_Y(\cdot)$, x , $\delta\sigma_U$, sample design and sample size, design-based estimators of the second difference on the left hand side of (2.3) will tend to be more stable with larger c . The data analysis of Section 6 below used $c = 2$. For that specific dataset, moderate changes in c did not have a severe effect on the resulting estimates of $F_Y(\cdot)$ and associated quantiles. Additional study of appropriate choices of c , or of quadratic-regression and other estimators of $F^{(2)}(\cdot)$ mentioned in Section 2.3, would be of interest for future work.

4.3. Quantiles

A direct estimator of the true-observation quantile $x_{F\gamma}$ is

$$\tilde{x}_{F\gamma} \stackrel{\text{def}}{=} \inf\{x : \tilde{F}_Y(x) \geq \gamma\} = \tilde{F}_Y^{-1}(\gamma), \quad (4.4)$$

say. Also, a direct extension of the Woodruff (1952) method leads to the approximate $(1 - \alpha)100\%$ confidence interval,

$$(\tilde{x}_{F\gamma L}, \tilde{x}_{F\gamma U}) = [\tilde{F}_Y^{-1}\{\tilde{F}_{YL}(\tilde{x}_{F\gamma})\}, \tilde{F}_Y^{-1}\{\tilde{F}_{YU}(\tilde{x}_{F\gamma})\}]. \quad (4.5)$$

In addition, direct substitution of $\hat{H}_{X\delta}(\cdot)$ and $\hat{\sigma}_{U\delta}$ into expression (3.5) produces an estimator $\hat{\Delta}(x)$, say, of the shift function $\Delta(x)$ from expression (3.4). An associated confidence interval is

$$[\hat{\Delta}_L(x), \hat{\Delta}_U(x)] = \hat{\Delta}(x) \pm z_{1-\alpha/2} [\hat{V}\{\hat{\Delta}(x)\}]^{1/2}, \quad (4.6)$$

where the variance estimator $\hat{V}\{\hat{\Delta}(x)\}$ follows directly from standard linearization arguments.

Finally, in the present work we will use $\hat{\Delta}(x)$ as a diagnostic to explore the magnitude of measurement error effects at various points x . However, Luo (1996) and Luo and Stokes (1997) showed that one can also use $\Delta(x)$ as the basis for direct estimation of the true-value distribution function $F_Y(\cdot)$. Specifically, consider some estimators $H_{X\delta}^*(\cdot)$ and $\Delta^*(\cdot)$, say. Then the relation $F_Y(x) = H_{X\delta}\{x + \Delta(x)\}$ leads to an alternative estimator $F_Y^*(x) = H_{X\delta}^*\{x + \Delta^*(x)\}$; see Luo (1996) and Luo and Stokes (1997) for detailed discussion of $F_Y^*(\cdot)$.

5. Sensitivity Analyses

As noted in Section 4.1, the measurement-error variance estimator $\hat{\sigma}_{U\delta}^2$ generally is based on some form of auxiliary data, e.g., replicate measurements. Similar comments apply to related identifying information used in previous literature, e.g., Stefanski and Bay (1996) and Nusser, Carrquiry, Dodd and Fuller (1996). However, in some practical cases it is difficult or impossible to obtain this information. See, e.g., Biemer (1995, p.146). In addition, even for cases in which replicate measurements are available, there can be a substantial burden associated with data cleaning and the calculation of $\hat{\sigma}_{U\delta}^2$ and the corresponding terms of $\hat{V}_{\nu H\sigma}(x)$.

Thus, it is useful to complement the adjusted estimation methods of Section 4 with some simple sensitivity-analysis tools. These tools use limited information (e.g., a range of possible measurement error variances, based on equipment specifications or related previous studies) to identify cases for which measurement error is of greatest concern in estimation of $F_Y(\cdot)$ or related quantiles. For these identified cases, one may proceed with a deeper (and potentially more time consuming) study, e.g., collection of replicate measurements and a more detailed subsequent measurement error analysis as in Section 4.

Specifically, assume that external sources of information have provided a data analyst with a range of plausible values $[m_L, m_U]$, say, for the measurement error variance $\sigma_{U\delta}^2$. For a given element $m \in [m_L, m_U]$, define the point estimator $F_Y^*(x)$ and the confidence interval $(F_{Y_L}^*(x), F_{Y_U}^*(x))$ by expressions (4.2) and (4.3) with $\hat{\sigma}_{U\delta}^2$ replaced by m and the corresponding elements of $\hat{V}_{\nu H\sigma}(x)$ set equal to zero. In addition, define the quantile estimator $x_{F\gamma}^*$ and confidence

interval $(x_{F\gamma L}^*, x_{F\gamma U}^*)$ through expressions (4.4) and (4.5), with the terms $\tilde{F}_Y(\cdot)$, $\tilde{F}_Y^{-1}\{\tilde{F}_{YL}(\tilde{x}_{F\gamma})\}$ and $\tilde{F}_Y^{-1}\{\tilde{F}_{YU}(\tilde{x}_{F\gamma})\}$ replaced accordingly. Similarly, define the shift function point estimator $\Delta^*(x)$ and confidence interval $[\Delta_L^*(x), \Delta_U^*(x)]$ through corresponding modifications of expressions (3.5) and (4.6).

6. Application to NHANES III Data

6.1. Hemoglobin measurements in NHANES III

An application of the Section 5 sensitivity analyses arose with data from measurements X of hemoglobin (measured in grams per deciliter) collected in the U.S. Third National Health and Nutrition Examination Survey (NHANES III). See National Center for Health Statistics (1996) for a detailed discussion of the NHANES III design and for a description of appropriate design-based analysis methods. For the current discussion, it suffices to note that NHANES III is a large-scale survey based on a stratified multistage design with 49 strata, two primary sample units (roughly equivalent to counties) selected per stratum, and additional subsampling used to select secondary sample units (roughly equivalent to city blocks), households and individual persons within selected households. Point estimates and variance estimates for the observed-value distribution function $H_{X\delta}(\cdot)$ were computed using standard methods. For example, point estimators $\hat{H}_{X\delta}(x)$ were standard weighted ratios, with weights computed from inverses of selection probabilities, with poststratification adjustments to account for nonresponse and related issues. In addition, $\hat{V}\{\hat{H}_{X\delta}(x)\}$ was computed from standard variance estimation formulas for a weighted sample ratio under a complex design as in, e.g., Cochran (1977).

Dallman Looker, Johnson and Carroll (1996) give a detailed discussion of hemoglobin and other iron status measures in NHANES III. For the current discussion, it suffices to note that very low levels of hemoglobin are associated with anemia, a serious health problem. Consequently, epidemiologists have strong interest in the quantiles of the distribution of hemoglobin in certain demographic groups. In addition, it is important to have estimates of low-iron prevalence rates as determined by some specific standard cutoffs. For example, for white women aged 20-49, a hemoglobin level of less than 12 grams per deciliter is an indicator of anemia. Thus, epidemiologists will have specific interest in estimation of the low-hemoglobin prevalence rate equal to $F_Y(12)$.

All hemoglobin measurements, including those collected by NHANES III, are believed to contain some amount of measurement error. In particular, preliminary work with some related NHANES III data indicated that the measurement error for hemoglobin had a variance in the range $[m_1, m_2] = [0.15, 0.30]$. In keeping with the small-error ideas of Sections 2 and 3, note that the square

roots of these possible variances (e.g., $(m_1)^{1/2} = 0.387$ or $(m_2)^{1/2} = 0.548$) are fairly small relative to overall variability in the X_δ observations. For example, $\hat{H}_{X_\delta}(10) = 0.011$ and $\hat{H}_{X_\delta}(15) = 0.964$; i.e., the effective support for hemoglobin measurements is roughly the interval $[10, 15]$.

6.2. Prevalence rates for low hemoglobin levels, and related quantiles

The current analysis focused on hemoglobin levels for white females aged 20-49. Table 1 presents point estimates and approximate 95% confidence intervals for $F_Y(12)$, the true proportion of persons in this population with hemoglobin less than 12 grams per deciliter. The first row gives the unadjusted estimates, i.e., estimates computed under the assumption that $\sigma_{U_\delta}^2 = 0$. The remaining rows report estimates for $\sigma_{U_\delta}^2 = 0.15(0.05)0.30$, in keeping with the sensitivity analyses outlined in Section 5. Also, as discussed at the end of Section 4.2, the present analysis used the multiplier $c = 2$ in the computation of expressions (4.1) and (4.2) and associated quantiles, shift functions and confidence bounds.

Table 1. Point estimates $F_Y^*(12)$ and approximate 95% confidence intervals $(F_{Y_L}^*(12), F_{Y_U}^*(12))$ for the prevalence of low hemoglobin. NHANES III data for white women aged 20-49.

	$F_Y^*(12)$	$(F_{Y_L}^*(12), F_{Y_U}^*(12))$
Unadjusted ($m = 0$)	0.1321	(0.1123, 0.1520)
$m = 0.15$	0.1195	(0.0989, 0.1402)
$m = 0.20$	0.1162	(0.0953, 0.1371)
$m = 0.25$	0.1121	(0.0910, 0.1331)
$m = 0.30$	0.1010	(0.0888, 0.1312)

Table 2. Point estimates $x_{F,0.05}^*$ and approximate 95% confidence intervals $(x_{F,0.05,L}^*, x_{F,0.05,U}^*)$ for the 0.05 quantile of hemoglobin. NHANES III data for white women aged 20-49.

	$x_{F,0.05}^*$	$(x_{F,0.05,L}^*, x_{F,0.05,U}^*)$
Unadjusted ($m = 0$)	11.27	(11.16, 11.42)
$m = 0.15$	11.42	(11.24, 11.54)
$m = 0.20$	11.47	(11.26, 11.58)
$m = 0.25$	11.51	(11.32, 11.62)
$m = 0.30$	11.53	(11.35, 11.64)

Note especially that for $\sigma_{U_\delta}^2 = 0.30$, the adjusted confidence intervals fall entirely below the unadjusted point estimate. This implies that measurement error may have led to substantial inflation in the unadjusted prevalence rate. Table 2 presents related point estimates and confidence intervals for $x_{F,0.05}$, the fifth

percentile of the population hemoglobin distribution. The practical implications of this sensitivity analysis are qualitatively similar to those for Table 1.

To examine the quantile issue in further detail, Figure 1 presents a shift function plot for $\sigma_{U\delta}^2 = 0.25$. The center curve represents the point estimates $\Delta^*(x)$, while the upper and lower curves represent the upper and lower pointwise 95% confidence bounds $\Delta_U^*(x)$ and $\Delta_L^*(x)$. Note especially that for x between 10 and 12, the shift function is fairly constant, at about -0.10 . In addition, note the relatively tight confidence bounds for $\Delta(x)$ with $x \leq 13$. As one would expect, shift-function plots using other values of $\sigma_{U\delta}^2$ showed more (less) pronounced shifts for larger (smaller) values of $\sigma_{U\delta}^2$; otherwise, the plots were qualitatively similar to Figure 1.

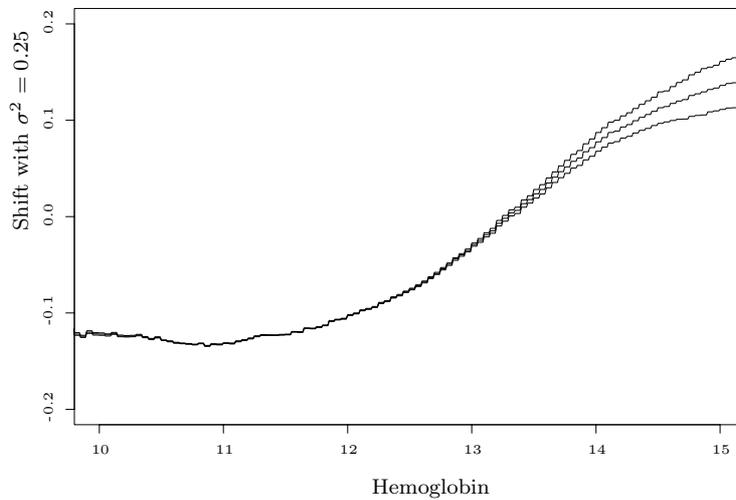


Figure 1. Plot of shift-function point estimates $\Delta^*(x)$ and approximate 95% confidence bounds $[\Delta_L^*(x), \Delta_U^*(x)]$ for hemoglobin levels with $m = 0.25$. NHANES III data for white females aged 20-49.

7. Discussion

Taken together, this paper, Stefanski and Bay (1996) and Nusser Carriquiry, Dodd and Fuller (1996) indicate that the choice of a specific measurement-error adjustment method for estimation of $F_Y(\cdot)$ depends heavily on the characteristics of the measurement errors, and on available information and analytic resources. For example, Sections 2 and 4 of the present paper indicate that if the measurement error variance is relatively small, then one can obtain relatively simple adjusted point estimators that depend only on the unadjusted estimator $\hat{H}_{X\delta}(\cdot)$ and an estimator $\hat{\sigma}_{U\delta}^2$ of the measurement error variance. This adjusted estimator does not require the assumption of normal measurement errors. In addition, the

adjusted estimators and diagnostics in Sections 4 and 5 are relatively simple to implement. This is important because some previously proposed transformation and adjustment methods generally require substantial analytic effort. See, e.g., Fuller (1995, p.127). Also, the proposed simple adjustments extend directly to construction of adjusted quantile estimators, the associated shift function $\Delta(x)$ and the sensitivity analyses discussed in Sections 5 and 6.

On the other hand, the methods in Stefanski and Bay (1996) and Nusser Carriquiry, Dodd and Fuller (1996) also have strengths. For example, if measurement errors are believed to be normal, as well as relatively small, then extensions of expression (9) in Stefanski and Bay (1996) indicate that simulation-extrapolation methods can lead to more refined adjustments (e.g., with remainder terms of order $O(\delta^6)$ instead of the order $O(\delta^3)$ or $O(\delta^4)$ remainders discussed in Sections 2 and 3 here for nonnormal errors). In addition, if the variances of measurement errors and true values are of roughly the same magnitude, then the small-error approximation approaches in this paper and in Stefanski and Bay (1996) may be problematic. For such cases, the transformation approach in Nusser Carriquiry, Dodd and Fuller (1996) is attractive because under its stated conditions it provides additional structure through the assumption that both the transformed measurement errors and the transformed true values arise from normal distributions.

Acknowledgement

The author thanks V.L. Parsons, the editor, an associate editor and an anonymous referee for helpful comments on an earlier version of this paper; S. L. Stokes for a helpful discussion of Luo (1996) and Luo and Stokes (1997); V. Burt, C. Johnson, A. Looker, and V.L. Parsons for helpful comments on the NHANES III application in Section 6; and S.R. Lee and S.Y. Heo for carrying out some of the computing work for Section 6. This research was supported in part by the U.S. National Center for Health Statistics. The views expressed here are those of the author and do not necessarily represent the policies of the U.S. National Center for Health Statistics.

Appendix. Proofs of Results 2.1-2.3 and 3.1-3.2, and Three Related Lemmas

This appendix contains proofs of the main results of Sections 2 and 3. It also states and proves the related technical Lemmas A.1 through A.3, which are used in the proof of Result 3.1.

Proof of Result 2.1. Given condition (C.1) and model (2.1), routine convolution arguments (e.g., Chung (1974), p.144) show that for a given real number x ,

the distribution function of X_δ is

$$H_{X_\delta}(x) = \int_{-\infty}^{\infty} F_Y(x - \delta t) dG_U(t). \tag{A.1}$$

Routine Taylor-expansion arguments for $F_Y(\cdot)$ show that under condition (C.2),

$$F_Y(x - \delta t) = F_Y(x) + F_Y^{(1)}(x)(-\delta t) + 2^{-1}F_Y^{(2)}(x)(-\delta t)^2 + 6^{-1}F_Y^{(3)}(x^*)(-\delta t)^3,$$

where x^* is a point in the interval $[x - \delta t, x]$. Thus, expression (A.1) equals

$$\begin{aligned} & \int_{-\infty}^{\infty} [F_Y(x) + F_Y^{(1)}(x)(-\delta t) + 2^{-1}F_Y^{(2)}(x)(\delta t)^2 + (6)^{-1}F_Y^{(3)}(x^*)(\delta t)^3] dG_U(t) \\ &= F_Y(x) \int_{-\infty}^{\infty} dG_U(t) + 2^{-1}F_Y^{(2)}(x) \int_{-\infty}^{\infty} (\delta t)^2 dG_U(t) \\ & \quad + \int_{-\infty}^{\infty} (6)^{-1}F_Y^{(3)}(x^*)(-\delta t)^3 dG_U(t) = F_Y(x) + 2^{-1}F_Y^{(2)}(x)\delta^2\sigma_U^2 + \delta^3 R_3(x), \end{aligned}$$

where $R_3(x)$ is a uniformly bounded term involving the bounded derivative $F_Y^{(3)}(\cdot)$ and the bounded third moment associated with $G_U(\cdot)$; and where the integration results follow from the fact that U has a mean equal to zero. Expression (2.2) follows immediately from the boundedness of $R_3(x)$.

Proof of Result 2.2. Under conditions (C.1) and (C.2), standard Taylor expansion arguments show that

$$F_Y(x + c\delta\sigma_U) = F_Y(x) + F_Y^{(1)}(x)c\delta\sigma_U + 2^{-1}F_Y^{(2)}(x)(c\delta\sigma_U)^2 + 6^{-1}F_Y^{(3)}(x^{**})(c\delta\sigma_U)^3, \tag{A.2}$$

and

$$F_Y(x - c\delta\sigma_U) = F_Y(x) + F_Y^{(1)}(x)(-c\delta\sigma_U) + 2^{-1}F_Y^{(2)}(x)(c\delta\sigma_U)^2 + 6^{-1}F_Y^{(3)}(x^{***})(-c\delta\sigma_U)^3, \tag{A.3}$$

where x^{**} and x^{***} are contained in the intervals $[x, x + c\delta\sigma_U]$ and $[x - c\delta\sigma_U, x]$, respectively. Also, due to the uniform bounding condition (C.2), $F_Y^{(3)}(x^{**})(c\delta\sigma_U)^3$ and $F_Y^{(3)}(x^{***})(c\delta\sigma_U)^3$ are both of order $O(\delta^3)$. Addition of expressions (A.2) and (A.3) then shows that

$$F_Y(x + c\delta\sigma_U) + F_Y(x - c\delta\sigma_U) - 2F_Y(x) = F_Y^{(2)}(x)(c\delta\sigma_U)^2 + O(\delta^3). \tag{A.4}$$

Condition (C.2) and an argument similar to that for expression (A.4) show that

$$F_Y^{(2)}(x + c\delta\sigma_U) + F_Y^{(2)}(x - c\delta\sigma_U) - 2F_Y^{(2)}(x) = O(\delta^2). \tag{A.5}$$

Three applications of Result 2.1, and expressions (A.4) and (A.5), show that

$$\begin{aligned} & H_{X\delta}(x + c\delta\sigma_U) + H_{X\delta}(x - c\delta\sigma_U) - 2H_{X\delta}(x) \\ &= F_Y(x + c\delta\sigma_U) + F_Y(x - c\delta\sigma_U) - 2F_Y(x) \\ &\quad + 2^{-1}\{F_Y^{(2)}(x + c\delta\sigma_U) + F_{X\delta}^{(2)}(x - c\delta\sigma_U) - 2F_{X\delta}^{(2)}(x)\}\delta^2\sigma_U^2 + O(\delta^3) \\ &= F_Y(x + c\delta\sigma_U) + F_Y(x - c\delta\sigma_U) - 2F_Y(x) + O(\delta^3) \\ &= F_Y^{(2)}(x)(c\delta\sigma_U)^2 + O(\delta^3), \end{aligned}$$

and expression (2.3) follows. Expression (2.4) then follows directly from expressions (2.2) and (2.3).

Proof of Result 2.3. Under the symmetry condition (2.5), expression (A.1) equals

$$\int_0^\infty \{F_Y(x + \delta t) + F_Y(x - \delta t)\}dG_U(t). \tag{A.6}$$

Also, Taylor-expansion arguments for $F_Y(\cdot)$ show that under under the stated derivative conditions,

$$\begin{aligned} & F_Y(x + \delta t) + F_Y(x - \delta t) \\ &= F_Y(x) + F_Y^{(1)}(x)(\delta t) + 2^{-1}F_Y^{(2)}(x)(\delta t)^2 + 6^{-1}F_Y^{(3)}(x)(\delta t)^3 \\ &\quad + (24)^{-1}F_Y^{(4)}(x^{****})(\delta t)^4 + F_Y(x) + F_Y^{(1)}(x)(-\delta t) + 2^{-1}F_Y^{(2)}(x)(-\delta t)^2 \\ &\quad + 6^{-1}F_Y^{(3)}(x)(-\delta t)^3 + (24)^{-1}F_Y^{(4)}(x^{*****})(-\delta t)^4 \\ &= 2F_Y(x) + F_Y^{(2)}(x)(\delta t)^2 + (24)^{-1}\{F_Y^{(4)}(x^{****}) + F_Y^{(4)}(x^{*****})\}(\delta t)^4, \end{aligned}$$

where x^{****} and x^{*****} are points in the intervals $[x, x + \delta t]$ and $[x - \delta t, x]$, respectively. Thus, expression (A.6) equals

$$\begin{aligned} & \int_0^\infty [2F_Y(x) + F_Y^{(2)}(x)(\delta t)^2 + (24)^{-1}\{F_Y^{(4)}(x^{****}) + F_Y^{(4)}(x^{*****})\}(\delta t)^4]dG_U(t) \\ &= 2F_Y(x) \int_0^\infty dG_U(t) + F_Y^{(2)}(x) \int_0^\infty (\delta t)^2 dG_U(t) \\ &\quad + \int_0^\infty (24)^{-1}\{F_Y^{(4)}(x^{****}) + F_Y^{(4)}(x^{*****})\}(\delta t)^4 dG_U(t) \\ &= F_Y(x) + 2^{-1}F_Y^{(2)}(x)\delta^2\sigma_U^2 + \delta^4 R_4(x), \end{aligned}$$

where $R_4(x)$ is a uniformly bounded term involving the bounded derivative $F_Y^{(4)}(\cdot)$ and the bounded fourth moment associated with $G_U(\cdot)$; and where the final integration results follow from the fact that $G_U(\cdot)$ is a symmetric distribution function. Expression (2.6) then follows immediately from the boundedness

of $R_4(x)$. Expression (2.7) then follows from arguments that are almost identical to those for Result 2.2, but with third-order Taylor expansions and order expressions replaced by corresponding fourth-order terms.

Lemma A.1. *Assume conditions (C.1) through (C.4). Then the following statements hold.*

- (a) *For any $\gamma \in (0, 1)$, there exist positive real numbers d_γ and $C_{1\gamma}$ such that $F_Y^{(1)}(x) > C_{1\gamma}$ for all x in the closed interval $[x_{F\gamma} - d_\gamma, x_{F\gamma} + d_\gamma]$.*
- (b) *For any $\gamma \in (0, 1)$ and for sufficiently small $\delta > 0$, $x_{H\delta\gamma} \in [x_{F\gamma} - d_\gamma, x_{F\gamma} + d_\gamma]$.*

Proof of Lemma A.1. Part (a) follows immediately from the continuity of $F_Y(x)$ and condition (C.4). To establish part (b), note first that a review of the proof of Result 2.1 shows that there exists some real number $C_{2\gamma} > 0$ such that $|H_{X\delta}(x_{F\gamma} - d_\gamma) - F_Y(x_{F\gamma} - d_\gamma)| \leq C_{2\gamma}\delta^2$ for all δ . Also, part (a) and a routine integration argument show that $F_Y(x_{F\gamma}) \geq F_Y(x_{F\gamma} - d_\gamma) + d_\gamma C_{1\gamma}$. Moreover, for sufficiently small δ , $d_\gamma C_{1\gamma} > C_{2\gamma}\delta^2$ and thus $H_{X\delta}(x_{F\gamma} - d_\gamma) \leq F(x_{F\gamma} - d_\gamma) + |H_{X\delta}(x_{F\gamma} - d_\gamma) - F_Y(x_{F\gamma} - d_\gamma)| \leq F_Y(x_{F\gamma} - d_\gamma) + C_{2\gamma}\delta^2 \leq F_Y(x_{F\gamma} - d_\gamma) + d_\gamma C_{1\gamma} \leq F_Y(x_{F\gamma}) = \gamma$, where the final equality follows from the definition of $x_{F\gamma}$ and the absolute continuity of $F_y(\cdot)$. Similar arguments show that $H_{X\delta}(x_{F\gamma} + d_\gamma) \geq \gamma$ for sufficiently small δ . Part (b) then follows immediately from the definition of $x_{H\delta\gamma}$.

Lemma A.2. *Assume conditions (C.1) through (C.4). Then for any $\gamma \in (0, 1)$ there exists $C_{3\gamma} > 0$ such that $|x_{H\delta\gamma} - x_{F\gamma}| \leq C_{3\gamma}\delta^2$.*

Proof of Lemma A.2. By condition (C.2) and the mean value theorem,

$$F_Y(x_{H\delta\gamma}) - F_Y(x_{F\gamma}) = F_Y^{(1)}(x^*)(x_{H\delta\gamma} - x_{F\gamma}), \tag{A.7}$$

where x^* is contained in the interval with endpoints $x_{H\delta\gamma}$ and $x_{F\gamma}$. Moreover, Lemma A.1 implies that for sufficiently small δ , $x_{H\delta\gamma} \in [x_{F\gamma} - d_\gamma, x_{F\gamma} + d_\gamma]$, so x^* is contained in the same interval, and thus $F_Y^{(1)}(x^*) > C_{1\gamma}$. Then by expression (A.7) $|x_{H\delta\gamma} - x_{F\gamma}| \leq (C_{1\gamma})^{-1}|F_Y(x_{H\delta\gamma}) - F_Y(x_{F\gamma})|$ for sufficiently small δ . In addition, a review of the proof of Result 2.1 shows that due to the uniform boundedness of the first through third derivatives of $F_Y(\cdot)$, there exists a positive real number C_4 such that $|F_Y(x_{H\delta\gamma}) - F_Y(x_{F\gamma})| \leq C_4\delta^2$ for all δ . Lemma A.2 then follows immediately, with $C_{3\gamma} = (C_{1\gamma})^{-1}C_4$.

Lemma A.3. *Assume conditions (C.1) through (C.4). Then for any $\gamma \in (0, 1)$ there exists $C_{5\gamma} > 0$ such that $|F_Y^{(2)}(x_{H\delta\gamma}) - F_Y^{(2)}(x_{F\gamma})| \leq C_{5\gamma}\delta^2$.*

Proof of Lemma A.3. By condition (C.2) and the Mean Value Theorem, $F_Y^{(2)}(x_{H\delta\gamma}) - F_Y^{(2)}(x_{F\gamma}) = F_Y^{(3)}(x^{**})(x_{H\delta\gamma} - x_{F\gamma})$ where x^{**} is contained in the

interval with endpoints $x_{H\delta\gamma}$ and $x_{F\gamma}$. Lemma A.3 then follows from Lemma A.2 and the uniform boundedness of $F^{(3)}(\cdot)$ provided in condition (C.2).

Proof of Result 3.1. Condition (C.2) and a standard Taylor expansion show that

$$F_Y(x_{H\delta\gamma}) - F_Y(x_{F\gamma}) = F_Y^{(1)}(x_{F\gamma})(x_{H\delta\gamma} - x_{F\gamma}) + 2^{-1}F_Y^{(2)}(x^*)(x_{H\delta\gamma} - x_{F\gamma})^2, \quad (\text{A.8})$$

where x^* is a point in the interval bounded by $x_{H\delta\gamma}$ and $x_{F\gamma}$. An application of Lemma A.2, along with the uniform boundedness of $F^{(2)}(\cdot)$ provided by condition (C.2), shows that $F_Y^{(2)}(x^*)(x_{H\delta\gamma} - x_{F\gamma})^2 = O(\delta^4)$. Routine algebra and expression (A.8) then imply that

$$x_{H\delta\gamma} - x_{F\gamma} = \{F_Y^{(1)}(x_{F\gamma})\}^{-1}\{F_Y(x_{H\delta\gamma}) - F_Y(x_{F\gamma})\} + O(\delta^4). \quad (\text{A.9})$$

Also, $F_Y(x_{F\gamma}) = \gamma = H_{X\delta}(x_{H\delta\gamma})$, so

$$F_Y(x_{H\delta\gamma}) - F_Y(x_{F\gamma}) = F_Y(x_{H\delta\gamma}) - H_{X\delta}(x_{H\delta\gamma}) = -2^{-1}F_Y^{(2)}(x_{H\delta\gamma})\delta^2\sigma_U^2 + O(\delta^3), \quad (\text{A.10})$$

where the final equality follows from Result 2.1. Result 3.1 then follows from expressions (A.9) and (A.10), Lemma A.3 and routine algebra.

Proof of Result 3.2. Routine Taylor-expansion arguments show that under condition (C.2),

$$\begin{aligned} F_Y^{(1)}(x_{F\gamma}) &= (2c\delta\sigma_U)^{-1}\{F_Y(x_{F\gamma} + c\delta\sigma_U) - F_Y(x_{F\gamma} - c\delta\sigma_U)\} + O(\delta) \\ &= (2c\delta\sigma_U)^{-1}\{H_{X\delta}(x_{F\gamma} + c\delta\sigma_U) - H_{X\delta}(x_{F\gamma} - c\delta\sigma_U)\} + O(\delta), \end{aligned} \quad (\text{A.11})$$

where the final equality follows from Result 2.1. In addition, two applications of the Mean Value Theorem and Lemma A.2 show that expression (A.11) equals

$$(2c\delta\sigma_U)^{-1}\{H_{X\delta}(x_{H\delta\gamma} + c\delta\sigma_U) - H_{X\delta}(x_{H\delta\gamma} - c\delta\sigma_U)\} + O(\delta). \quad (\text{A.12})$$

Similar arguments show that

$$F_Y^{(2)}(x_{F\gamma})(c\delta\sigma_U)^2 = \{H_{X\delta}(x_{H\delta\gamma} + c\delta\sigma_U) + H_{X\delta}(x_{H\delta\gamma} - c\delta\sigma_U) - 2H_{X\delta}(x_{H\delta\gamma})\} + O(\delta^4). \quad (\text{A.13})$$

Then by Result 3.1,

$$\begin{aligned} x_{F\gamma} &= x_{H\delta\gamma} + 2^{-1}\{F_Y^{(1)}(x_{F\gamma})\}^{-1}F_Y^{(2)}(x_{F\gamma})\delta^2\sigma_U^2 + O(\delta^3) \\ &= x_{H\delta\gamma} + 2^{-1}[(2c\delta\sigma_U)^{-1}\{H_{X\delta}(x_{H\delta\gamma} + c\delta\sigma_U) - H_{X\delta}(x_{H\delta\gamma} - c\delta\sigma_U)\} + O(\delta)]^{-1} \\ &\quad \times \{H_{X\delta}(x_{H\delta\gamma} + c\delta\sigma_U) + H_{X\delta}(x_{H\delta\gamma} - c\delta\sigma_U) - 2H_{X\delta}(x_{H\delta\gamma}) + O(\delta^4)\} \\ &= x_{H\delta\gamma} + 2^{-1}[(2c\delta\sigma_U)^{-1}\{H_{X\delta}(x_{H\delta\gamma} + c\delta\sigma_U) - H_{X\delta}(x_{H\delta\gamma} - c\delta\sigma_U)\}]^{-1} \\ &\quad \times \{H_{X\delta}(x_{H\delta\gamma} + c\delta\sigma_U) + H_{X\delta}(x_{H\delta\gamma} - c\delta\sigma_U) - 2H_{X\delta}(x_{H\delta\gamma})\} + O(\delta^3), \end{aligned}$$

where the second equality follows from expressions (A.12) and (A.13), and the final equality follows from additional routine Taylor expansion arguments.

References

- Bahadur, R. R. (1966). A note on quantiles in small samples. *Ann. Math. Statist.* **37**, 577-580.
- Biemer, P. P. (1995). Discussion of paper by W.A. Fuller. *Internat. Statist. Rev.* **63**, 143-147.
- Binder, D. A. (1983). On the variances of asymptotically normal estimators from complex surveys. *Internat. Statist. Rev.* **51**, 279-292.
- Box, G. E. P. and Cox, D. R. (1964). An analysis of transformations (with discussion). *J. Roy. Statist. Soc. Ser. B* **26**, 211-246.
- Carroll, R. J. and Stefanski, L. A. (1990). Approximate quasi-likelihood estimation in models with surrogate predictors. *J. Amer. Statist. Assoc.* **85**, 652-663.
- Carroll, R. J., Ruppert, D. and Stefanski, L. A. (1995). *Measurement Error in Nonlinear Models*. Chapman and Hall, New York.
- Chesher, A. (1991). The effect of measurement error. *Biometrika* **78**, 451-462.
- Chung, K. L. (1974). *A Course in Probability Theory*. Academic Press, New York.
- Dallman, P. R., Looker, A. C., Johnson, C. L. and Carroll, M. (1996). Influence of age on laboratory criteria for the diagnosis of iron deficiency anaemia and iron deficiency in infants and children. In *Iron Nutrition in Health and Disease* (Edited by L. Hallberg and N.-G. Asp), 65-74. John Libby and Co., London.
- Doksum, K. A. and Sievers, G. L. (1976). Plotting with confidence: graphical comparisons of two populations. *Biometrika* **63**, 421-434.
- Eckert, R. S., Carroll, R. J. and Wang, N. (1997). Transformations to additivity in measurement error models. *Biometrics* **53**, 262-272.
- Eddington, A. S. (1913). On a formula for correcting statistics for the effects of a known probable error of observation. *Monthly Notices Roy. Astronom. Soc.* **73**, 359-360.
- Francisco, C. A. and Fuller, W. A. (1991). Quantile estimation with a complex survey design. *Ann. Statist.* **19**, 454-469.
- Fuller, W. A. (1975). Regression analysis for sample survey. *Sankhyā Ser. C* **37**, 117-132.
- Fuller, W. A. (1995). Estimation in the presence of measurement error. *Internat. Statist. Rev.* **63**, 121-141.
- Gaffey, W. R. (1959). A consistent estimator of a component of a convolution. *Ann. Math. Statist.* **30**, 198-205.
- Korn, E. L. and Graubard, B. I. (1998). Variance estimation for superpopulation parameters. Unpublished manuscript, National Cancer Institute, Bethesda, Maryland.
- Kovar, Rao, J. N. K. and Wu, C. F. J. (1988). Bootstrap and other methods to measure errors in survey estimates. *Canadian J. Statist.* **16**, S, 25-45.
- Luo, M. (1996). Estimation of the CDF of a finite population in the presence of measurement error. Unpublished Ph.D. dissertation, Department of Management Science and Information Systems, University of Texas at Austin.
- Luo, M. and Stokes, L. (1997). Estimation of the CDF of a finite population in the presence of measurement error. Paper presented at the International Conference on Current Topics in Survey Sampling, Lincoln, Nebraska, October 24-26, 1997.
- Mee, R. W., Owen, D. B. and Shyu, J.-C. (1986). Confidence bounds for misclassification probabilities based on data subject to measurement error. *J. Qual. Tech.* **18**, 29-40.
- National Center for Health Statistics (1996). NHANES III Reference Manuals and Reports. CD-ROM, GPO 017-022-01358-4, United States Government Printing Office, Washington, DC.
- Nusser, S. M., Carriquiry, A. L., Dodd, K. W. and Fuller, W. A. (1996). A semiparametric transformation approach to estimating usual daily intake distributions. *J. Amer. Statist. Assoc.* **91**, 1440-1449.

- Pollard, H. (1953). Distribution functions containing a Gaussian factor. *Proc. Amer. Math. Soc.* **4**, 578-582.
- Rao, J. N. K., Kovar, J. G. and Mantel, H. J. (1990). On estimating distribution functions from survey data using auxiliary information. *Biometrika* **77**, 365-375.
- Shao, J. (1996). Resampling methods in sample surveys (with discussion). *Statistics* **27**, 203-254.
- Shao, J. and Wu, C. F. J. (1992). Asymptotic properties of the balanced repeated replication method for sample quantiles. *Ann. Statist.* **20**, 1571-1593.
- Stefanski, L. A. and Bay, J. M. (1996). Simulation extrapolation deconvolution of finite population cumulative distribution function estimators. *Biometrika* **83**, 407-417.
- Woodroffe, M. (1975). *Probability with Applications*. McGraw-Hill, New York.
- Woodruff, R. S. (1952). Confidence intervals for medians and other position measures. *J. Amer. Statist. Assoc.* **47**, 635-646.
- Zhong, C. X. B. (1997). Empirical likelihood inference in the presence of measurement error. *Proceedings of the Section on Survey Research Methods, American Statistical Association*.

Department of Statistics, Texas A&M University, College Station, TX 77843-3143, U.S.A.
E-mail: jeltinge@stat.tamu.edu

(Received May 1997; accepted July 1998)