*J. R. Statist. Soc.* A (2001)
**164**, *Part* 3, *pp.* 505–516

# Estimating reference ranges from stratified two-stage samples

Mary Lou Thompson, Steven D. Edland, Laura E. Gibbons and Susan M. McCurry

*University of Washington, Seattle, USA*

[Received May 1999. Final revision January 2001]

**Summary.** We present methodology for estimating age-specific reference ranges by using data from two-stage samples. On the basis of the information obtained in the first stage, the initial sample is stratified and random subsamples are drawn from each stratum, where the selection probabilities in this second-stage sampling may be different across strata in the population. The variable for which the reference ranges are to be established is measured at the second phase. The approach involves maximum likelihood estimation of the parameters of the age-specific distributions and separate estimation of the population stratum probabilities. These are combined to yield estimates of the quantiles of interest. The issue of variance estimation for the estimated quantiles is also addressed. The methodology is applied to the estimation of reference ranges for a cognitive test score in a study of non-demented older Japanese-Americans.

*Keywords*: Cognitive test; Model assessment; Reference ranges; Two-stage sample

## 1. Introduction

Much attention has been focused in the statistics literature over the last decade on methods for estimating age- or time-dependent reference ranges or percentiles. A variety of parametric (Cole, 1988; Thompson and Theron, 1990; Altman, 1993; Royston and Wright, 1998), semiparametric (Cole and Green, 1992; Heagerty and Pepe, 1999) and nonparametric (Healy *et al.*, 1988; Pan *et al.*, 1990; Rossiter, 1991) approaches have been considered. There is also a separate but related body of work on so-called regression quantiles (see, for example, Koenker and Bassett (1978), He (1997) and Yu and Jones (1998)), where regression methods are applied directly to the percentiles of interest (as opposed to indirectly, via modelling of the distribution of the variable of interest).

This present application concerns the estimation of age-specific percentiles from stratified two-stage sampling, where the sampling fractions vary across strata. This kind of sampling occurs in epidemiological studies where some measures of interest which can be obtained easily or cheaply are measured in the first phase (where they may be used as a screening instrument) and in the second-stage sample (which may be selected on the basis of the screening scores) other variables which may be difficult to obtain are ascertained on a stratified subset of the original sample (Hand, 1987; Flanders and Greenland, 1991; Zhao and Lipsitz, 1992; Pickles *et al.*, 1995). For instance, a more definitive disease verification assessment may be carried out on the subsample at the second stage. In the context of psychiatric research, the verification of disease stage can include the collection of large amounts of

*Address for correspondence*: Mary Lou Thompson, Department of Biostatistics, University of Washington, Box 357232, Seattle, WA 98195, USA.
E-mail: mlt@u.washington.edu

 0964–1998/01/164505

cognitive, functional and affective test data from diverse populations. These data could be useful in the development of instrument norms for age and ethnic groups that have traditionally been under-represented in normative studies. However, this use is contingent on an appropriate adjustment for the sampling design. Lipsitz *et al.* (1997) considered an approach to quantile regression in the context of a longitudinal study with non-random drop-out which is effectively a similar setting.

In this paper, we consider the extension of a parametric approach to estimating centile curves to the setting where only second-stage information is available on some of the variables of interest but where the goal is to estimate percentiles for a general population. We describe the type of study design being considered and introduce the notation in Section 2. The example of application is introduced in Section 3. For given stratum probabilities, we develop the approach to estimating percentiles of interest in Section 4. We then consider the question of estimating the stratum probabilities in the context of two-stage sampling. Section 5 contains an example of application to the cognitive abilities screening instrument (CASI) (Teng *et al.*, 1994) in a population of older Japanese-Americans as part of the Kame project, a National Institute on Aging prospective study in Seattle–King County, Washington (Graves *et al.*, 1996). In Section 6, we consider the (often-overlooked) issue of the precision of the centile curves and Section 7 contains a discussion.

## 2. Design and notation

There are many variations on two-stage (or two-phase) designs, particularly in the setting of case–control or case–cohort studies (see, for example Zhao and Lipsitz (1992)). We consider a design where the clinical assessment and definitive determination of the disease occur only at the second stage of the study on a subsample, stratified within the original sample according to information collected at the first stage.

We assume that, at the first stage, a random sample is selected from the population of interest and, for each individual in the sample, a variable $Z = 1, 2, \ldots, I$ indicating stratum membership is measured, in addition to covariates, $\mathbf{X}_c$. Typically $Z$ is some screening indicator of disease status. On the basis of $Z$ and possibly also $\mathbf{X}_c$, a stratified subsample of the original sample is obtained at the second stage, where the outcome of interest, $Y$, and other covariates, $\mathbf{X}_u$, are measured on a random sample of individuals within each stratum. We denote the second-stage sample by $S_2$ and the first-stage sample (on which there is only partial information, namely on $Z$ and $\mathbf{X}_c$) by $S_1$. We assume that, in the second-stage sampling, $Y_{ij}$ has been measured on $j = 1, 2, \ldots, N_i$ individuals in each of the $i = 1, 2, \ldots, I$ strata. $\mathbf{X}_{ij}$ denotes the covariate information that is available on these individuals from both phases of sampling. The population probability of an individual with covariate pattern $\mathbf{x}$ being in the $i$th stratum is denoted by $p_i(\mathbf{x})$.

In studies of dementia, for instance, $Z$ may be based on the result of a cognitive screening test and $\mathbf{X}_c$ might be the age of the subject. Within strata defined by $Z$ (and possibly further depending on $\mathbf{X}_c$), random subsamples are selected at the second stage for further cognitive testing and the determination of the status of dementia and other covariates (such as the level of education). In our context, we are interested in establishing reference ranges for the cognitive tests ($Y$), possibly as a function of covariates ($\mathbf{X}$). The specific study design that we consider in our application is described below.

## 3. The Kame study

As part of a longitudinal study to investigate the incidence of dementia in a Japanese-

American population, 1993 individuals with at least 50% Japanese heritage and aged 65 years and older were given an initial assessment to screen for prevalent cognitive impairment. The screening instrument was the CASI which has a score ranging from 0 to 100. Dementia was not assessed at this stage. On the basis of their CASI scores and their ages, the population was divided into three CASI strata (CASI $< 81$, $81 \leqslant$ CASI $< 87$, CASI $\geqslant 87$) and five age groups (65–69, 70–74, 75–79, 80–84 and 85 years and older). Individuals were sampled from within these groups for subsequent follow-up. There was oversampling of lower CASI strata and the older age groups who were at a greatest risk of being diagnosed with dementia, with 384 individuals being sampled in total. In the framework outlined above, $Z$ represents the three CASI strata and $X_c$ is age, regarded as a continuous covariate. At the second stage, individuals were reassessed with a standard neuropsychological test battery, including a re-evaluation of their current CASI score. Interviews with patients and proxy informants and a physical and neurological examination were also obtained. Completed evaluations were discussed at consensus meetings attended by a psychologist, a geriatrician, a neurologist, an epidemiologist and research staff, who were blinded to the screening CASI score. On this basis, 235 of the 384 individuals sampled at the second stage were determined to be non-demented, 196 of whom completed the second CASI questionnaire. Of the remaining 39 non-demented individuals who were sampled, their other available neuropsychological information and evaluation by the physician were sufficient to determine the dementia status. Three of these individuals refused to complete the CASI questionnaire and, for the others, the questionnaire was completed but the results were not regarded as valid, in most instances because of problems of hearing. The modelling of the stratum-specific distributions in Section 5.1 is based on the 196 non-demented individuals for whom CASI scores are available and we assume that no bias is introduced by the missing data.

In the framework above, $Y$ represents an item in the neuropsychological test battery and $\mathbf{X}_u$ is the dementia status and other covariates such as educational level. A more detailed description of the study design and the rationale behind it may be found in Graves *et al.* (1996). Of interest here is the estimation of reference ranges for the various neuropsychological test outcomes, particularly in the non-demented population.

## 4. Weighted centile estimation

We focus first on modelling the distribution of $Y$, given $i$ (stratum) and $\mathbf{X}$. Here we consider only $S_2$, the stratified subsample of individuals for whom information is available at both the first and the second stage of sampling. We assume that

$$g(Y)|i, \mathbf{X} \sim \mathcal{N}(f_1(i, \mathbf{X}), f_2(i, \mathbf{X})) \tag{1}$$

where $g$ is a monotone transformation, depending on unknown parameters and possibly on $i$ (stratum) and $\mathbf{X}$. The joint density of $W_{ij} = g(Y_{ij})$, given $\mathbf{X}_{ij}$, is then given by

$$\prod_{i=1}^{I} \prod_{j=1}^{N_i} p_i(\mathbf{X}_{ij})\, \phi(W_{ij}; f_1(i, \mathbf{X}_{ij}), f_2(i, \mathbf{X}_{ij})) \tag{2}$$

where $\phi(w; f_1, f_2)$ is the normal density function with mean $f_1$ and standard deviation $f_2$, evaluated at $w$.

The $k$th percentile of $Y$, given $\mathbf{X}$, is then $g^{-1}(c_k)$ where $c_k$ is the solution to

$$k = \sum_{i=1}^{I} p_i(\mathbf{X})\, \Phi(c_k;\, f_1(i, \mathbf{X}),\, f_2(i, \mathbf{X})) \tag{3}$$

and where $\Phi$ is the normal distribution function. Estimation of the quantiles of $Y$ requires the estimation of the parameters of $g$, $f_1$ and $f_2$, and of the probability that an individual is in the $i$th stratum, given covariate pattern $\mathbf{X} = (\mathbf{X}_c,\, \mathbf{X}_u)$. Given estimates of the components of equation (3), the non-linear equation for $c_k$ can then be solved by standard iterative techniques, such as Newton's method.

Given $p_i(\mathbf{X})$ and parametric forms for $g$, $f_1$ and $f_2$, the parameters may be estimated by maximizing expression (2). Note that the maximum likelihood solution for the parameters of $g$, $f_1$ and $f_2$ is functionally independent of the stratum probabilities $p_i$. The variance–covariance matrix of the parameter estimates may be obtained from the estimated information in the usual way.

We next turn our attention to the estimation of the stratum probabilities $p_i(\mathbf{x})$. At this point we consider information from $S_1$ and $S_2$. We assume a polytomous logistic regression model for the stratum probabilities, given $\mathbf{X}$:

$$\text{logit}\{P(Z \leqslant i|\mathbf{X})\} = \alpha_{0i} + \boldsymbol{\alpha}_1\mathbf{X} \qquad i = 1, 2, \ldots, I-1. \tag{4}$$

Approaches to the statistical analysis of two-stage studies have been considered for the special case of two strata (generally, diseased and non-diseased) by several researchers (see, for example, Roberts *et al*. (1987), Cain and Breslow (1988) and Zhao and Lipsitz (1992)). We follow the approach of Reilly and Pepe (1995), who developed a mean score method for general regression models with missing covariate data.

If a model for the conditional distribution of $\mathbf{X}_u$, given $\mathbf{X}_c$, were specified, then $P(Z|\mathbf{X}_c)$ can be derived from equation (4) by taking the expectation of $P(Z|\mathbf{X})$ over the distribution of $\mathbf{X}_u|\mathbf{X}_c$. A likelihood approach to the estimation of $\boldsymbol{\alpha}$ would then involve the maximization of

$$\sum_{k\in S_2} \log\{P(Z_k|\mathbf{X}_k)\} + \sum_{k\notin S_2} \log\{P(Z_k|\mathbf{X}_{ck})\}.$$

An EM approach would consider the maximization of

$$\sum_{k\in S_2} \log\{P(Z_k|\mathbf{X}_k)\} + \sum_{k\notin S_2} E[\log\{P(Z_k|\mathbf{X}_k)\}]$$

where the expectation is with respect to $\mathbf{X}_{uk}|\mathbf{X}_{ck}$, holding $\boldsymbol{\alpha}$ at its previous value.

For discretized $\mathbf{X}_{ck}$, Reilly and Pepe (1995) proposed estimating the second term by

$$\hat{E}[\log\{P(Z_k|\mathbf{X}_k)\}] = \frac{1}{m_{2k}} \sum_{s\in S_2} \log\{P(Z_s|\mathbf{X}_s)\}\, I_k(s)$$

where $I_k(s) = 1$ when $\mathbf{X}_{cs} = \mathbf{X}_{ck}$, $Z_s = Z_k$ and $m_{ik} = \Sigma_{j\in S_i} I_k(j)$. The estimated log-likelihood then becomes

$$\sum_{k\in S_2} \frac{m_{1k}}{m_{2k}} \log\{P(Z_k|\mathbf{X}_k)\},$$

which is the complete-data likelihood with inverse probability weighting (Horvitz and Thompson, 1952). Estimation can hence be implemented by using standard software and a weighting function.

Reilly and Pepe (1995) also derived general expressions for the variance–covariance matrix

of $\hat{\alpha}$ resulting from the mean score approach which can be adapted to the special case of a polytomous logistic model. The estimation requires an adjustment to the variance that would have been obtained if the second-stage information had been available on all the subjects. In essence, a penalty is applied for the incomplete sampling.

## 5. Application to the Kame study

### 5.1. Fitting the stratum-specific distributions

We present here methodology for estimating reference ranges for any of the neurocognitive measures that were obtained at the second stage of this study. The approach is illustrated by the application to the CASI score in non-demented individuals. Summary statistics for second-stage CASI scores for the non-demented in each of the strata are presented in Table 1, where $m_1$, $m_2$ and $n$ represent respectively the number of individuals at the base-line, the number of individuals in $S_2$ and the number of non-demented individuals for whom CASI scores are available in the second-stage sample.

In the framework developed above, the variable of interest, $Y$, is the CASI score in the *non-demented* population. The CASI score, like any cognitive screening measure, cannot be collected on individuals with severe dementia. In addition, the CASI score may be used clinically to follow the progression of disease and in this setting it is useful to establish the decline that is to be expected in non-demented subjects. Several additional covariates are of potential interest in this context (education, primary language and gender), but for ease of exposition we confine our attention here to age $X_c$ and dementia status $X_u$. $X_c$ is measured at the first stage of the study on all individuals and $X_u$ only at the second on the stratified subsample. Given age, there are $I = 3$ strata, defined by three base-line CASI score

**Table 1.** Summary of the CASI second-stage non-demented sample†

| *Screening* | $m_1, m_2, n$ | $Q_1$ | *Median* | $Q_3$ |
|---|---|---|---|---|
| *Stratum 1* | | | | |
| $65 \leqslant$ Age $< 70$ | 31, 22, 14 | 79.5 | 84.5 | 87.0 |
| $70 \leqslant$ Age $< 75$ | 52, 42, 27 | 76.0 | 80.5 | 84.5 |
| $75 \leqslant$ Age $< 80$ | 57, 49, 23 | 77.2 | 82.6 | 86.0 |
| $80 \leqslant$ Age $< 85$ | 43, 38, 10 | 77.3 | 78.5 | 86.5 |
| Age $\geqslant 85$ | 122, 106, 17 | 65.5 | 72.2 | 77.5 |
| *Stratum 2* | | | | |
| $65 \leqslant$ Age $< 70$ | 70, 14, 12 | 87.5 | 91.5 | 93.6 |
| $70 \leqslant$ Age $< 75$ | 84, 18, 15 | 85.0 | 87.5 | 90.5 |
| $75 \leqslant$ Age $< 80$ | 80, 23, 20 | 81.2 | 86.2 | 88.9 |
| $80 \leqslant$ Age $< 85$ | 26, 7, 6 | 82.5 | 86.7 | 88.0 |
| Age $\geqslant 85$ | 19, 16, 10 | 78.5 | 84.5 | 87.5 |
| *Stratum 3* | | | | |
| $65 \leqslant$ Age $< 70$ | 640, 3, 3 | 92.0 | 93.5 | 97.5 |
| $70 \leqslant$ Age $< 75$ | 506, 15, 14 | 90.0 | 92.4 | 96.5 |
| $75 \leqslant$ Age $< 80$ | 179, 8, 7 | 84.5 | 90.5 | 93.0 |
| $80 \leqslant$ Age $< 85$ | 66, 7, 5 | 89.5 | 91.0 | 94.5 |
| Age $\geqslant 85$ | 18, 16, 13 | 82.5 | 84.8 | 88.5 |

†$m_1$, $m_2$ and $n$ are the number of individuals sampled at the base-line and at the second stage and the number of non-demented individuals with CASI scores at the second stage respectively.

categories. The second-stage sampling was carried out within three CASI strata and five age groups. We model age as a continuous covariate $X_c$, which supersedes the dependence on the age groups in the stratum centile estimation. The goal of the analysis is to estimate age-specific reference ranges for the variable CASI for the non-demented population. Although the methodology developed in Section 4 would in principle allow a simultaneous estimation of the parameters of the CASI distributions in the demented and non-demented populations, we base this part of the modelling and estimation solely on the sample of non-demented individuals (and hence $X_u$ is dropped from this part of the analysis). Because of the relatively small size of the sample, the percentiles of interest were restricted to the quartiles and the median.

The first step in this data analysis is the identification of a monotone function $g$, which will transform the CASI scores to normality. Our approach here followed that of Cole (1988) and explored the range of Box–Cox power transformations. The exploratory analysis revealed no obvious dependence of the power transform on age although polynomial dependence on age was considered in the modelling. Exploratory plots of grouped means and standard deviations were also used to guide the choice of the forms of $f_1$ and $f_2$. Here again, a range of polynomial models in $X_c$, including interactions with the levels of $Z$, were considered. Likelihood ratio tests were used to compare nested models.

Let $Z_1$ and $Z_2$ be dummy variables for the two upper base-line CASI strata. The final fitted model was of the form

$$g(Y) = (Y^{\lambda} - 1)/\lambda,$$
$$f_1 = \beta_1 + \beta_2 Z_1 + \beta_3 Z_2 + \beta_4 X_c + \beta_5 X_c Z_2,$$
$$f_2 = \exp(\beta_6 + \beta_7 Z_2 + \beta_8 Z_2 X_c).$$

In the modelling, the mirror image of $Y$, i.e. $100 - Y$, was found to be more amenable to a normalizing power transformation. The maximum likelihood estimation was carried out using the International Mathematical and Statistical Libraries' routine DUMIDH (International Mathematical and Statistical Libraries, 1987). The fitted model for the mean, for instance, has a different intercept for each of the three strata and a different slope with age for the highest base-line CASI stratum. The fitted model for the standard deviation is constant (over age) for the lower two strata and changes linearly with age in the upper stratum.

The point estimates of the fitted models were

$$\hat{g}(y) = \{(100 - y)^{0.15} - 1\}/0.15,$$
$$\hat{f_1} = 1.12 - 0.649 Z_1 - 3.68 Z_2 + 0.0341 X_c + 0.0313 X_c Z_2,$$
$$\hat{f_2} = \exp(-0.607 + 3.34 Z_2 - 0.0403 Z_2 X_c).$$

Note that the fitted mean level of $Y$ decreases with increasing age and, not surprisingly, increases with increased base-line CASI score.

### 5.2. Estimating the stratum probabilities

The fitted models above estimate a distribution for $Y$ in the non-demented population for any given age and base-line stratum. The aim of the analysis, however, is to estimate reference ranges for the general non-demented population. Given $p_i(\mathbf{x})$, this involves solving equation (3) above. To estimate smoothly varying percentiles with age, it is necessary (and biologically plausible) for the probability of being in the $i$th stratum to change smoothly with age. We fitted the following polytomous logistic model:

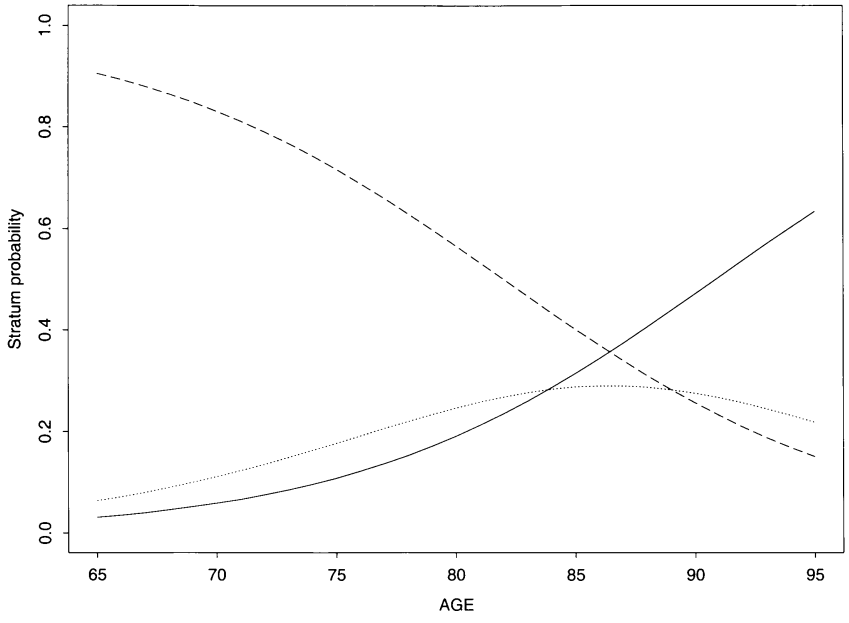


**Fig. 1.** Fitted stratum probabilities: ———, stratum 1; ·····, stratum 2, - - - - -, stratum 3

$$\text{logit}\{P(Z \leqslant i|\mathbf{X})\} = \alpha_{0i} + \alpha_1 X_c + \alpha_2 X_u \qquad i = 1, 2$$

where $X_c$ and $X_u$ are age and dementia status with $X_u = 1$ for demented individuals and $X_u = 0$ otherwise. Models with higher order terms in $X_c$ and $X_c$, $X_u$ interactions were also considered but did not contribute significantly to the fit of the model. As outlined in Section 4, estimation from the two-stage data followed the mean score approach of Reilly and Pepe (1995) which was implemented by using SAS procedure `LOGISTIC` (SAS Institute, 1990) to fit the polytomous logistic regression model, with weights equal to the ratio of the number of first-stage to second-stage individuals in each stratum.

The fitted model is of the form

$$\text{logit}\{\hat{P}(Z = i|X_c, X_u)\} = \hat{\alpha}_{0i} + 0.133X_c + 1.77X_u \qquad i = 1, 2$$

where $\hat{\alpha}_{01} = -12.09$, $\hat{\alpha}_{02} = -10.90$ and the probabilities for the third stratum are obtained by subtraction. Note that the estimated probability of lower stratum membership (i.e. lower base-line CASI score) is higher for demented and older individuals. We are interested in the fitted probabilities for the non-demented population, corresponding to the case $X_u = 0$. The fitted stratum probabilities are shown in Fig. 1. Given the estimates of $p_i(\mathbf{x})$, we can estimate the percentiles for the general non-demented population, by using equation (3). Fig. 2 shows the estimated quartiles and median for the non-demented population as a function of age.

## 6. Model assessment

The percentiles estimated above assume a Gaussian distribution on the transformed scale. Fig. 3 shows normal probability plots for the standardized residuals from the fit in three age categories (under 70, 70–79 and 80 years and older), which indicate no lack of fit, apart from

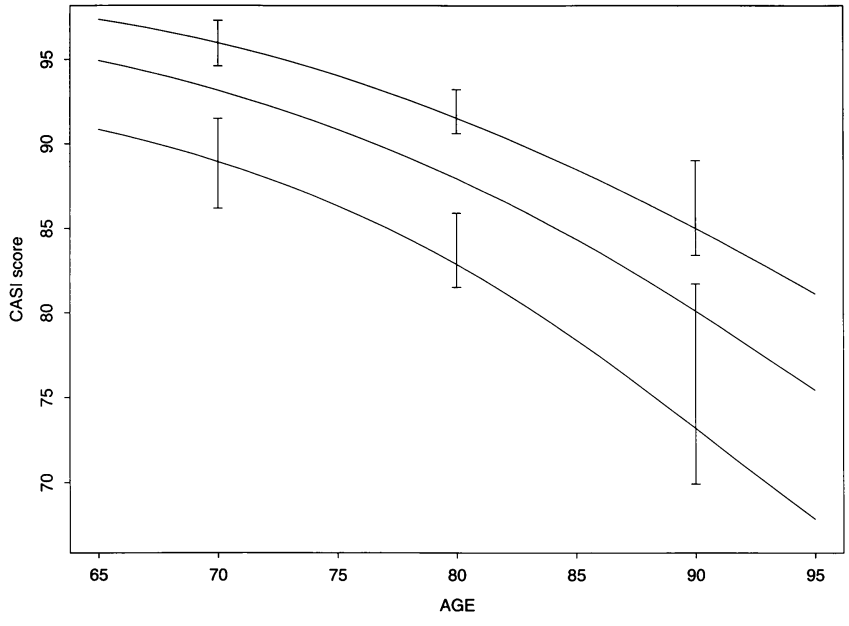


**Fig. 2.** Estimated median and quartiles

a single outlier in the first plot. This observation corresponds to a mentally retarded 69-year-old individual with a CASI score of 40 (the estimated lower quartile of the CASI score at age 69 years is 89.4). Fig. 4 shows the fitted stratum-specific and overall CASI score densities for the 75- and 85-years-old groups.

The precision of the fitted percentiles should also be considered, particularly with such a relatively small data set. Our approach to evaluating the precision was as follows. The maximum likelihood and mean score procedures described in Section 4 provide estimates of the covariance matrices of the fitted parameters of the stratum-specific distributions of $Y$ and of the stratum probabilities $p_i(\mathbf{x})$. Assuming a joint multivariate normal distribution for each set of parameters with mean and variance–covariance set equal to the estimated values (with independence between the estimates from the maximum likelihood and mean score components), we simulated 1000 samples of the parameter estimates from these distributions and, on the basis of these, calculated the corresponding percentile estimates from equation (3). The resulting pointwise 95% confidence bands for the upper and lower quartiles at ages 70, 80 and 90 years based on a consideration of both sources of variability are also shown in Fig. 2.

## 7. Discussion

Two-stage sampling is used in epidemiological research generally and in studies of Alzheimer's disease in particular. In the latter case, the primary goal is typically the establishment of the incidence of dementia and its association with relevant covariates. A secondary interest, however, is the use of the large body of neuropsychological test scores that are collected to provide guidelines on 'normal' ranges in the non-demented population. The estimation of these reference ranges is complicated by the stratified sampling design. The approach described here allows the estimation of such reference ranges, incorporating data from both stages of the sample. Although in this example reference ranges were estimated for a variable

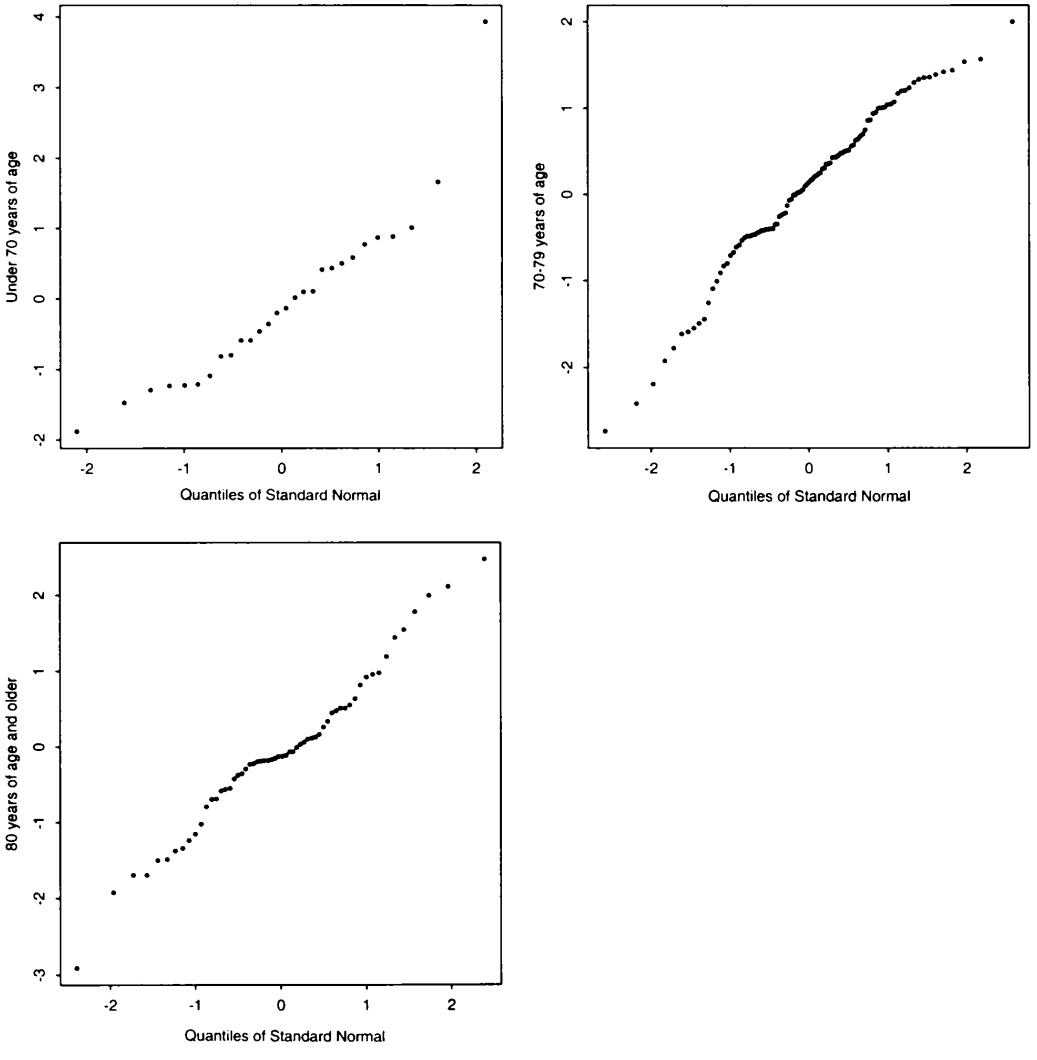


**Fig. 3.** Normal probability plots for standardized residuals

(the CASI score) that was measured at the first and second stage, the methodology applies generally to any variable that is measured at the second stage.

Many cognitive abilities decline with advancing age and mental status examinations such as the CASI include items assessing functions that are sensitive to age effects. As expected, CASI scores decrease with age. It is also to be expected that there is greater variability with increasing age as here factors such as significant health problems compound age effects in the elderly. Many studies of aging have reported larger variances for older adult samples than for younger groups (see, for example, LaRue (1992)).

A variety of approaches to the estimation of reference ranges has been proposed in the literature. We have considered here the particular setting where variables for which reference ranges are to be constructed are measured at the second stage of a two-stage study in which

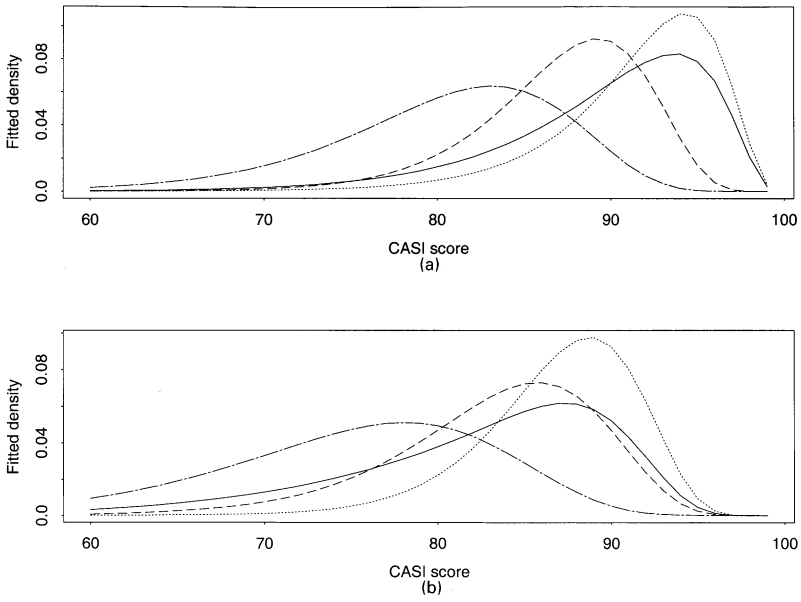


**Fig. 4.** Fitted densities at ages (a) 75 and (b) 85 years: ——— overall; — · — , stratum 1; - - - - -, stratum 2; ·····, stratum 3

the sampling proportions of different strata are different from the population proportions. Although we have developed a particular parametric approach to estimating the stratum-specific cross-sectional distributions, the approach developed here could be applied (using the equivalent of equation (3)) under any method of centile estimation that produces a stratum-specific distribution function as a function of age and covariates. A more flexible alternative to the polynomial models considered here for $g$, $f_1$ and $f_2$ might be the use of fractional polynomials (see, for example, Royston and Wright (1998)). The weighted quantile regression approach suggested by Lipsitz *et al.* (1997) would allow a direct estimation of the population level quantiles without a consideration of the stratum-specific distributions. In modelling the underlying distribution of the variable of interest, we acknowledge the inherent ordering of quantiles. It is possible, for instance, that quantile estimates from the quantile regression approach may cross. In addition, an exploratory analysis to guide the choice of an appropriate functional form for the models is likely to be more accessible at the stratum-specific level.

Similarly, other approaches to estimating the stratum probabilities might be considered. We chose to apply the methodology developed by Reilly and Pepe (1995) because it can be implemented by using standard software and because estimates of the variances of the estimates are readily obtained. The variability in the estimation of the weights did not contribute substantially to the variability in the centile estimates.

The assessment of model diagnostics and variability should always play an important part in any estimation of reference ranges and yet it is often neglected. The approximate 95% confidence bounds displayed in Fig. 2 show clearly that, for this small data set, the quartile estimates at the extremes of the age range are associated with considerable variability, particularly the lower quartile. The lack of precision demonstrated in Fig. 2 suggests expanding the second-stage sample. The optimal sampling strategies for two-stage studies considered by Reilly (1996) might be implemented in this regard.

The effects of age, education and language or cultural biases on psychometric test score performances are well known (Heaton *et al*., 1996; Geisinger, 1994). However, normative studies are expensive and difficult to conduct, and few data are available that describe how elderly, undereducated or culturally diverse individuals perform on standard tests of cognitive, functional or affective status. Studies that do exist are often based on very small sample sizes, clinical populations or convenience samples. The present paper has demonstrated that data collected during the screening and assessment phases of large epidemiological studies are one potential source of descriptive information on the test performance of individuals in under-represented populations. In the application presented here, our approach can also be used to examine the effects on test performance of other variables such as gender, education and language.

## Acknowledgements

The authors wish to acknowledge valuable comments by two reviewers and the Joint Editor. We also thank Marie Reilly and Margaret Pepe for helpful comments on earlier drafts. This research was funded in part by the Kame survey of dementia in the elderly Japanese-American population of King County, Washington (National Institute on Aging grant AG09769) and by the University of Washington Alzheimer's Disease Research Center (National Institute on Aging grant AG05136). The authors wish to acknowledge the Kame project's Principal Investigator, Dr Eric Larson, as well as other participating medical personnel, research staff and data support staff, without whom this study would not have been possible.

## References

Altman, D. G. (1993) Construction of age-related reference centiles using absolute residuals. *Statist. Med.*, **12**, 917–924.

Cain, K. C. and Breslow, N. E. (1988) Logistic regression analysis and efficient design for two-stage studies. *Am. J. Epidem.*, **128**, 1198–1206.

Cole, T. J. (1988) Fitting smoothed centile curves to reference data. *J. R. Statist. Soc.* A, **151**, 385–406.

Cole, T. J. and Green, P. J. (1992) Smoothing reference centile curves: the LMS method and penalized likelihood. *Statist. Med.*, **11**, 1305–1319.

Flanders, W. D. and Greenland, S. (1991) Analytic methods for two stage case-control studies and other stratified designs. *Statist. Med.*, **10**, 739–747.

Geisinger, K. F. (1994) Cross-cultural normative assessment: translation and adaptation issues influencing the normative interpretation of assessment instruments. *Psychol. Assessmnt*, **6**, 304–312.

Graves, A. B., Larson, E. B., Edland, S. D., Bowen, J. D., McCormick, W. C., McCurry, S. M., Rice, M. M., Wenzlow, A. and Uomoto, J. M. (1996) Prevalence of dementia and its subtypes in the Japanese American population of King County, Washington State: the Kame Project. *Am. J. Epidem.*, **144**, 760–771.

Hand, D. J. (1987) Screening *vs* prevalence estimation. *Appl. Statist.*, **36**, 1–7.

He, X. (1997) Quantile curves without crossing. *Am. Statistn*, **51**, 186–192.

Heagerty, P. J. and Pepe, M. S. (1999) Semiparametric estimation of regression quantiles with application to standardizing weight for height and age in US children. *Appl. Statist.*, **48**, 533–551.

Healy, M. J. R., Rasbash, J. and Yang, M. (1988) Distribution-free estimation of age-related centiles. *Ann. Hum. Biol.*, **15**, 17–22.

Heaton, R. K., Ryan, L., Grant, I. and Matthews, C. G. (1996) Demographic influences on neuropsychological test performance. In *Neuropsychological Assessment of Neuropsychiatric Disorders* (eds G. Igor and K. M. Adams), 2nd edn, pp. 141–163. New York: Oxford University Press.

Horvitz, D. G. and Thompson, D. J. (1952) A generalization of sampling without replacement from a finite universe. *J. Am. Statist. Ass.*, **47**, 663–685.

International Mathematical and Statistical Libraries (1987) *IMSL User's Manual: Math/Library*. Houston: International Mathematical and Statistical Libraries.

Koenker, R. and Bassett, G. (1978) Regression quantiles. *Econometrica*, **46**, 33–50.

LaRue, A. (1992) *Aging and Neuropsychological Assessment*. New York: Plenum.

Lipsitz, S. R., Fitzmaurice, G. M., Molenberghs, G. M. and Zhao, L. P. (1997) Quantile regression methods for longitudinal data with drop-outs: application to CD4 cell counts of patients infected with the human immunodeficiency virus. *Appl. Statist.*, **46**, 463–476.

Pan, H. Q., Goldstein, H. and Yang, Q. (1990) Nonparametric estimation of age-related centiles over wide age ranges. *Ann. Hum. Biol.*, **17**, 475–481.

Pickles, A., Dunn, G. and Vazquez-Barquero, J. L. (1995) Screening for stratification in two-phase epidemiological surveys. *Statist. Meth. Med. Res.*, **4**, 73–89.

Reilly, M. (1996) Optimal sampling strategies for two-stage studies. *Am. J. Epidem.*, **143**, 92–100.

Reilly, M. and Pepe, M. S. (1995) A mean score method for missing and auxiliary covariate data in regression models. *Biometrika*, **82**, 299–314.

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

Rossiter, J. E. (1991) Calculating centile curves using kernel density estimation methods with application to infant kidney lengths. *Statist. Med.*, **10**, 1693–1701.

Royston, P. and Wright, E. M. (1998) A method for estimating age-specific reference intervals ('normal ranges') based on fractional polynomials and exponential transformation. *J. R. Statist. Soc.* A, **161**, 79–101.

SAS Institute (1990) *SAS/STAT User's Guide, Version 6*. Cary: SAS Institute.

Teng, E. L., Hasegawa, K., Homma, A., Imai, Y., Larson, E. B., Graves, A. B., Sugimoto, K., Yamaguchi, T., Sasaki, H., Chiu, D. and White, L. R. (1994) The Cognitive Abilities Screening Instrument (CASI): a practical test for cross-cultural studies of dementia. *Int. Psychgeriatr.*, **6**, 45–58.

Thompson, M. L. and Theron, G. B. (1990) Maximum likelihood estimation of reference centiles. *Statist. Med.*, **9**, 539–548.

Yu, K. and Jones, M. C. (1998) Local linear quantile regression. *J. Am. Statist. Ass.*, **93**, 228–237.

Zhao, L. P. and Lipsitz, S. (1992) Designs and analysis of two-stage studies. *Statist. Med.*, **11**, 769–782.