*Public Health Nutrition: 6(4), 393–399* DOI: 10.1079/PHN2002446

# Misclassification associated with measurement error in the assessment of dietary intake

Carl de Moor[1,*], Tom Baranowski[2], Karen W Cullen[2] and Theresa Nicklas[2]
[1]Departments of Behavioral Science and Biostatistics, The University of Texas M.D. Anderson Cancer Center, 1515 Holcombe Boulevard – Box 243, Houston, TX 77030, USA: [2]Children's Nutrition Research Center, Department of Pediatrics, Baylor College of Medicine, Houston, TX, USA

Submitted 1 August 2002: Accepted 11 November 2002

## Abstract

*Objective:* Dietary assessment has been used for certification to receive food supplements or other nutrition services and to provide feedback for educational purposes. The proportion of individuals correctly certified as eligible is a function of the amount of error that exists in the dietary measures and the level of dietary intake used to establish eligibility. Whether individuals are correctly counselled to increase or decrease the consumption of selected foods or nutrients is a function of the same factors. It is not clear, however, what percentage of individuals would be correctly classified under what circumstances. The objective of this study is to demonstrate the extent to which measurement error and eligibility criteria affect the accuracy of classification.
*Design:* Hypothetical distributions of dietary intake were generated with varying degrees of measurement error. Different eligibility criteria were applied and the expected classification rates were determined using numerical methods.
*Setting and subjects:* Simulation study.
*Results:* Cut points of dietary intake at decreasing levels below the 50th percentile of true intake were associated with lower sensitivity and predictive value positive rates, but higher specificity and predictive value negative rates. The correct classification rates were lower when two cut points of dietary intake were used. Using a single cut point that was higher than the targeted true consumption resulted in higher sensitivity but lower predictive value positive, and lower specificity but higher predictive value negative.
*Conclusions:* Current methods of dietary assessment may not be reliable enough to attain acceptable levels of correct classification. Policy-makers and educators must consider how much misclassification error they are willing to accept and determine whether more intensive methods are necessary.

**Keywords**
Misclassification
Measurement error
Certification
Nutrition counselling

For some government food programmes, such as Women, Infants and Children (WIC), certification of a health or dietary need is required to receive nutrition service[1]. Given the large numbers of participants and limited staffing, only a brief dietary assessment lasting a few minutes is typically used as the basis for certification of dietary inadequacy. Similarly, in some nutrition education programmes, feedback or education is provided to the participant based on some form of dietary assessment[2,3].

Most of the common dietary assessment procedures have been used to characterise the average consumption of groups of people[4–6]. The high levels of error associated with these methods are usually accepted because the procedures are relatively inexpensive to administer to large numbers of individuals and the measurement errors are assumed to average out across the individuals, resulting in reasonably accurate group estimates. However, more reliable procedures are required to characterise the dietary consumption of individuals accurately. These generally involve more time-intensive multiple 24-hour dietary recalls, lengthy food-frequency questionnaires or one-week diet records. However, even these procedures may yield dietary estimates that contain substantial levels of measurement error. For example, correlations between food-frequency questionnaires and the standards against which they have been compared typically range from 0.5 to 0.6[7], indicating that approximately 75–64% of the variance in the estimates was error.

From the perspective of programme certification, measurement error has serious consequences because individuals who truly meet the dietary criteria for certification may not receive the proffered services, and

**Corresponding author:* Email cdemoor@mdanderson.org 

individuals who truly do not meet the dietary criteria may nevertheless receive them. The degree to which the certification process minimises classification errors determines how efficiently and successfully a programme allocates its resources. Similarly, measurement error for tailoring dietary feedback could provide erroneous information or encouragement (e.g. not encouraging those individuals who do not eat 5 portions of fruit and vegetables a day or encouraging those who already do).

The proportion of individuals who are correctly classified is partly a function of the amount of random error that exists in the dietary measures. Other important factors (ignoring bias) include the level of dietary consumption that is used to define eligibility and whether a single cut point or multiple cut points of dietary intake are used as the criterion. In combination, these factors can result in high rates of misclassification in all but the most favourable and perhaps unrealistic circumstances.

While the ideas on which these analyses are based are not new, we could find only one reference to the level of misclassification associated with these factors[8]. Thus, the primary contribution of this paper is to use accepted statistical procedures to estimate the expected levels of misclassification and identify levels of validity coefficients needed to minimise misclassification error.

## Methods

To illustrate the relationship between measurement error, the eligibility criterion and classification rates, we assumed that the measured dietary intake was a function only of the true dietary intake plus some level of random error, which can be written as

$$\text{Observed intake} = \text{true intake} + \text{error}.$$

In addition, we assumed that the level of error was constant and independent of the level of the true intake, and that the average error across all individuals in the population was zero. This type of error is called random error. In statistical terms, if the error in the observed intake were random error, the observed intake would be an unbiased estimate of the true intake, or, alternatively, the expected value of the observed intake in the population would be equal to the expected value of the true intake. If we assumed that the average level of error were less than or greater than zero, the observed intake would consistently under– or overestimate the true intake. This type of error is called systematic error or bias. Although systematic error would also have important implications for classification, it is possible (although difficult) to correct using calibration techniques[9]. For the purposes of this illustration, we ignored systematic error.

To simplify the presentation, we assumed that both the observed and true intakes were normally distributed with means $\mu$ and variances $\sigma^2$, respectively. The variance of the observed intake is then $\sigma^2_{\text{obs}} = \sigma^2_{\text{true}} + \sigma^2_{\text{error}}$. We can write the correlation between the observed intake and the true intake, denoted by $\rho$, using the standard product moment representation:

$$\rho = \text{covariance between observed and true intake}/(\sigma_{\text{observed intake}} \times \sigma_{\text{true intake}}).$$

The proportion of random error in the observed intake can be written in terms of the correlation between the observed and true intakes, in which the higher the level of random error, the lower the correlation between the observed and the true intakes. Specifically,

$$\begin{aligned}\text{Proportion error in observed intake}\\ &= \sigma^2_{\text{error}}/(\sigma^2_{\text{true}} + \sigma^2_{\text{error}})\\ &= (\sigma^2_{\text{obs}} - \sigma^2_{\text{true}})/\sigma^2_{\text{obs}} \text{ (because } \sigma^2_{\text{error}} = \sigma^2_{\text{obs}} - \sigma^2_{\text{true}})\\ &= 1 - (\sigma^2_{\text{true}}/\sigma^2_{\text{obs}})\\ &= 1 - \rho^2\end{aligned}$$

or, equivalently,

$$\rho = (1 - \text{proportion error in observed intake})^{1/2}.$$

The correlation between the observed and true intakes will be used in the discussion below to denote the level of measurement error in the observed intake, with a correlation of zero ($\rho = 0$) indicating that all of the variability in observed intake is due to error, i.e. the observed and true intakes are independent, and a correlation of one ($\rho = 1$) indicating that none of the variability in observed intake is due to error, i.e. the observed and true intakes are equal.

The relationship between the observed and true intakes is depicted graphically in Fig. 1. The diagonal line corresponds to the points where observed intake is identical to true intake. Points above the diagonal line represent observed intake values that are greater than the true intake, and points below the line represent observed intake values that are less than the true intake. The ellipse can be thought of as representing the area of highest

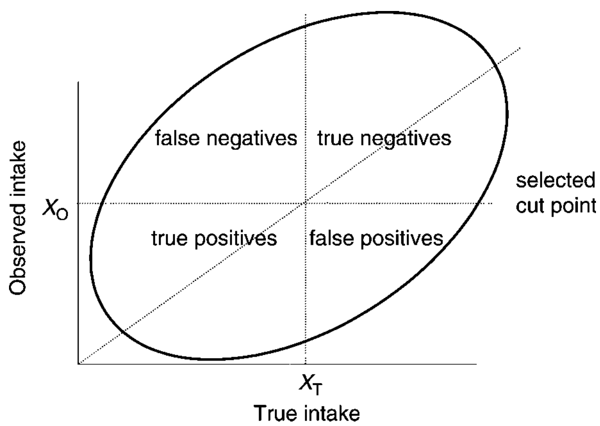


**Fig. 1** Identification of errors of assignment with a single cut point when the true values are known

concentration of the scatter plot of the observed and true intake values. The higher the correlation between the observed and true intakes, i.e. the less error in the observed data, the narrower the ellipse would be around the diagonal line. Conversely, the lower the correlation between the observed and true intakes, i.e. the more error in the observed intake, the wider the ellipse would be. $X_T$ represents a hypothetical cut point of true dietary intake that defines true eligibility, and the corresponding $X_O$ represents the same value based on observed intake (an alternative that we explore below sets the observed cut point $X_O$ to be greater than the targeted true intake cut point $X_T$). For example, an individual whose observed intake is less than $X_O$ would be *classified* as eligible to receive services, whereas an individual whose true intake is less than $X_T$ would be *truly* eligible to receive services. The lines associated with the cut points identify regions of different classification possibilities. For example, true positives are those individuals whose observed intake falls below the $X_O$ value and whose true intake is below $X_T$. Similarly, sensitivity is defined as the number of true positives divided by the number of individuals whose true intake is below $X_T$.

Dietary assessment in the past has been primarily concerned with identifying under-consumption as a risk for nutrient deficiency diseases[10]. With the onset of the recent national epidemic of obesity[11] and increasing attention on identifying optional levels of consumption of selected nutrients[1], there has been increasing concern with over-consumption. This has led to an interest in identifying mid-ranges that represent the desired or optimal levels of consumption. Therefore, the effects of eligibility criteria comprising two cut points of dietary intake are also of interest and discussed below.

For the analyses below, eligibility criteria that were based on a single cut point used the 10th, 25th, 33rd and 50th percentiles of the distribution of true intake ($X_T$). In these cases, the observed intake cut point was chosen to be equivalent to the targeted true intake cut point. Thus, a subject whose observed intake fell below the lowest 10th, 25th, 33rd or 50th percentile of the distribution of true intake would be classified as eligible, and an individual whose observed intake fell above these cut points would be classified as ineligible. The selected values were simply chosen to represent a wide range of potential cut points and were not obtained from actual programmes. For the analyses where eligibility was defined by two regions of dietary consumption, we used cut points corresponding to the 5th to 95th, 12.5th to 87.5th, 16.5th to 83.5th and 25th to 75th percentiles of true intake. These criteria defined eligibility as observed dietary consumption that fell in the outer 10, 25, 33 and 50% of the true intake of individuals, respectively, percentages which equalled those based on a single cut point. An individual whose observed intake fell outside the cut points was classified as eligible, and an individual whose observed intake fell between the cut points was classified as ineligible. Finally, we evaluated a scenario in which the cut point of the observed intake corresponded to the 20th percentile of true intake, when the targeted cut point was the 10th percentile of true intake.

For each scenario, the expected classification rates were given in terms of the sensitivity (SEN), predictive value positive (PVP), specificity (SPEC), predictive value negative (PVN) and accuracy (AC) of the classification criteria. SEN is the proportion of all truly eligible individuals who are classified as eligible; PVP is the proportion of all individuals classified as eligible who were truly eligible (true positives); SPEC is the proportion of truly ineligible individuals who are classified as ineligible; PVN is the proportion of all individuals classified as ineligible who were truly ineligible (true negatives); and AC is the proportion of all individuals who were correctly classified. To calculate the classification rates, the error model of dietary consumption presented above was used, with the assumption that the true intake and measurement error followed a joint bivariate normal distribution. The values of the classification rates were determined using numerical integration of the region of the joint distribution defined by the eligibility criteria. The classification results are depicted graphically for ease of interpretation.

## Results

In Figs 2, 3 and 4, we present the SEN/PVP, SPEC/PVN and AC for the scenarios in which a single cut point equivalent to the targeted true intake cut point was used to define eligibility. Because the observed cut point was equal to the targeted true intake cut point, SEN and PVP are equivalent, as are SPEC and PVN. Each of the lines in the figures represents the classification rates as a function of the correlation between the observed intake and the true intake. As can be seen in Fig. 2, PVP varied between a high of 1, which represents 100% SEN/PVP (true positives), and a low corresponding to the percentile used for the cut point. So for example, when the observed intake cut point was the 10th percentile of true intake, the lowest SEN/PVP value was 10% when all of the observed intake was error and thus independent of the true intake (i.e. correlation = 0). The curves indicated that the level of SEN/PVP was quite low for levels of correlation that have been reported between measures of dietary intake and dietary standards. For example, for a correlation between observed and true intakes of 0.50, the SEN/PVP rates varied from a low of 30% to a high of approximately 65%. Cut points based on lower percentiles resulted in a lower level of SEN/PVP than those based on a higher percentile (for cut points less than the 50th percentile). The results for SPEC and PVN revealed opposite patterns (Fig. 3). These values were higher than the SEN/PVP counterparts for all cut points except when the cut point was based on

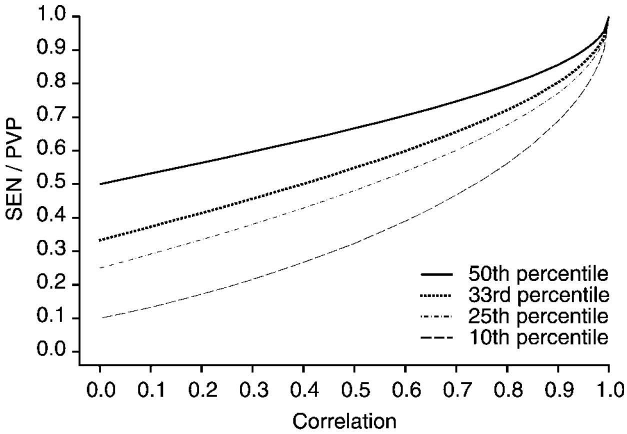


**Fig. 2** Sensitivity/predictive value positive (SEN/PVP) for four percentile cut points as the correlation between true and observed values increases

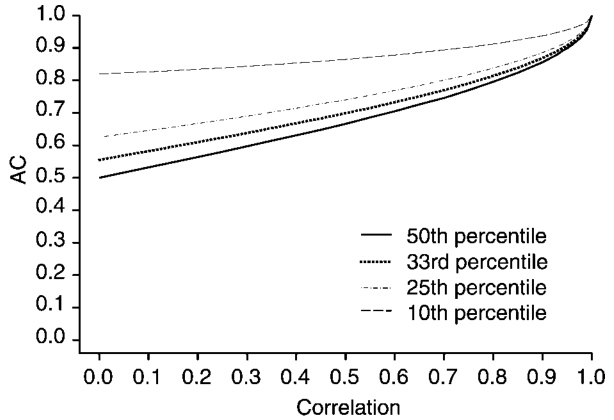


**Fig. 4** Accuracy (AC) for four percentile cut points as the correlation between true and observed values increases

the median, the point at which the SEN/PVP and SPEC/PVN were identical. The AC figures (Fig. 4) were between those of the SEN/PVP and SPEC/PVN and more closely reflected the SPEC/PVN rates due to the fact that the cut points presented here were all below the median.

Next, the classification rates for a single cut point were compared with those for two cut points, i.e. a range of values defined eligibility. The SEN/PVP rates in Fig. 5 are represented for the ranges corresponding to the 16.5th to 83.5th percentiles, where ineligibility was defined as intake in the outer 33% of all individuals. In this figure, the classification rates corresponding to the single cut point at the 33rd percentile (which was also presented in Figs 2 to 4 above) were also included. This allowed comparison of the classification rates when the lower 33% were used as the eligible group, versus the outer 33% beyond the 16.5th and 83.5th percentiles. The use of two regions of values had a moderately detrimental effect on the classification rates (figures corresponding to SPEC/PVN and AC are not presented owing to space limitations). Even when the correlation between the observed and true intakes was as high as 0.8, the SEN/PVP was only approximately 60%, indicating that roughly 40% of the subjects who were classified as eligible were in fact not eligible. The same pattern that appeared in the SEN/PVP plot were observed in SPEC/PVN, and as expected AC was thus substantially worse for the double than for the single cut point. Similar types of patterns were also observed when the other sets of double cut points were used (data also not presented).

Finally we present a scenario in which the observed intake cut point was set higher than the targeted true intake value. This may be done in practice by programme sponsors who want to ensure that as many of the truly eligible individuals are included as possible. We present the SEN and PVP for these results in Fig. 6. In this figure, the dashed line represents the SEN when the observed cut point is based on the 20th percentile and the targeted true intake is the 10th percentile, the solid line represents SEN/PVP when the observed cut point is equivalent to the targeted 10th percentile of true intake (also presented in Fig. 2) and the dotted line represents the PVP when the observed cut point is the 20th percentile and the targeted

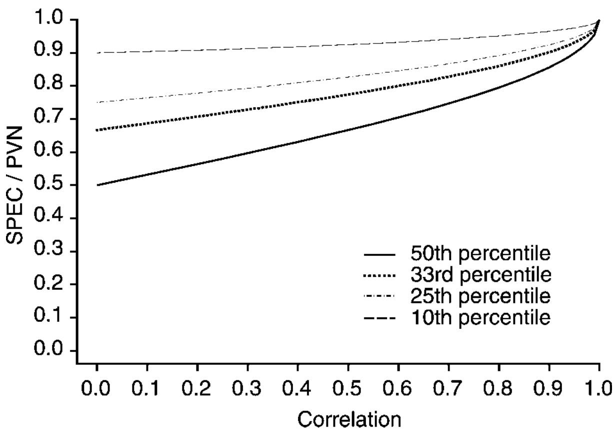


**Fig. 3** Specificity/predictive value negative (SPEC/PVN) for four percentile cut points as the correlation between true and observed values increases

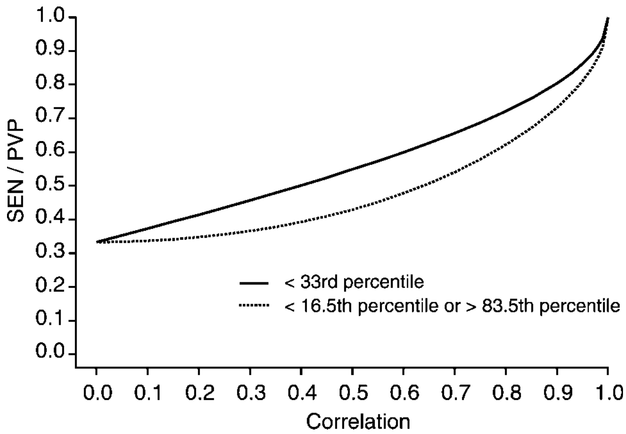


**Fig. 5** Sensitivity/predictive value positive (SEN/PVP) for a single cut point and double cut points, both certifying 33% of the population as eligible

https://doi.org/10.1079/PHN2002446 Published online by Cambridge University Press

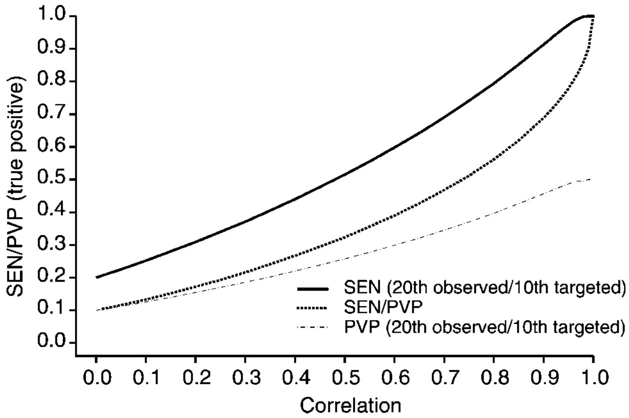


**Fig. 6** Sensitivity (SEN) and predictive value positive (PVP) for a single cut point set higher than the targeted value

true intake is the 10th percentile (in this case, because the observed and true intake cut points differ, the SEN and PVP are no longer equivalent). As expected, increasing the eligibility criteria to the 20th percentile resulted in an increase in the SEN above what occurred when the observed cut point was equivalent to the targeted true cut point. For example, when the correlation between observed and true intakes was as high as 0.80, the SEN increased from 56 to 79%, indicating that nearly four out of five subjects who were truly eligible were classified as eligible. However, at this same correlation, the PVP decreased to 40%, indicating that six out of 10 subjects who were classified as eligible are in fact ineligible. Similar patterns of change occurred for other scenarios in which we evaluated higher observed intake cut points than the targeted true intake cut points.

## Discussion

As the level of measurement error decreased, i.e. the correlation between true and observed intakes increased, the accuracy of a cut point in identifying true positives and true negatives increased. Cut points at decreasing levels below the 50th percentile of intake were associated with lower SEN/PVP (true positive) rates, but with higher PVN (true negative) rates. At levels of correlation between true and observed intakes that reflect values reported in the literature, e.g. 0.5 to 0.6, total accuracy varied from approximately 60 to 85%. Levels of accuracy of 90% or higher were attained only when the correlation was above 0.90, a situation not ordinarily attained in the dietary assessment literature. The classification rates were worse when double cut points were used to determine eligibility. Setting the observed cut point higher than the targeted true intake increased the SEN and decreased the PVP, resulting in an increase in the number of subjects who would receive services but not truly be eligible for them.

These findings only addressed issues of misclassification errors for different known levels of measurement error and for different classification criteria. They have focused on the problems inherent in very rapid or brief assessment and used the error levels associated with the more commonly used dietary measures as the upper bound of the correlation between observed and true intakes for comparison. The use of the correlation coefficient as the indicator of measurement error facilitated explication in terms of a commonly used measure of association. Alternative indicators could have been used as well. These findings did not address related problems of dietary assessment, including that there may be some incentive for programme participants to provide inaccurate estimates in order to qualify for services. In addition, at this time the classification criteria that likely maximise health outcomes may not be known precisely. Moreover, recently published research[12] indicates that the validity of some measures such as food-frequency questionnaires may be overstated, suggesting that the actual correlations between observed and true intakes may be even lower than currently appreciated. The misclassification ratios may be even higher if these were factored into the analyses. Finally, this research is based on a simple error model using common distributional assumptions. The relevance of these results to real-world observations, which often do not meet these assumptions, is nonetheless maintained as distributional violations can usually be corrected with simple transformations.

These findings have immediate application in at least two contexts: dietary assessment for certification for receiving nutrition services[13] and dietary assessment for educational feedback[2,3]. The WIC programme can be used as an example of dietary assessment for certification. While WIC gives priority to anthropometric, biochemical and clinical criteria, 43.5% of pregnant participants, 40.5% of breast-feeding participants and 47.6% of postpartum participants qualified for WIC services based on dietary risk[13]. Among children (1 to 4 years of age), 64.1% qualified based on dietary risk[13]. Although it is not clear from the records available what percentages of participants qualified based on dietary risk alone, dietary risk appeared to be the primary criterion for qualifying children. Thus, assuming a correlation between observed and true intakes of 0.5 for dietary assessment among children in WIC (which is probably at the higher levels), as many as 24% of the children classified as eligible are truly ineligible, and 43% who were classified as ineligible are truly eligible.

The WIC programme can also be used as an example of dietary assessment for nutrition education or counselling purposes[2]. (Dietary assessment in WIC can serve both certification and education purposes.) It is not clear what percentage of WIC contacts attempted to tailor education to the empirically documented level of consumption[2]. When a participant does not meet a particular Food Guide Pyramid cut point, the educator admonishes or persuades the participant to do so. Since the validity and reliability of these instruments would be the same for certification or education, a clear problem emerges in that the extent to which error occurred in dietary assessment, education tailored to that level of consumption has a high probability of providing the wrong advice. Thus, again assuming a correlation between observed and true intakes of 0.5 for dietary assessment, if the desired cut point for vegetables were 3 servings per day and a child is eating only 2, he has a 30% chance of being classified as ineligible and receiving incorrect counselling (assuming a standard deviation of dietary consumption of 1.9 serving per day for children aged 1 to 4 years obtained from the Continuing Survey of Food Intakes by Individuals, 1994–96).

Finally, these results also have implication for 'tailoring' nutrition education messages to psychosocial characteristics[3]. In this case, specific statements (items) from a psychosocial scale (e.g. outcome expectancies) are selected. Such statements usually have a five-category response (e.g. agree very much, agree, not sure, disagree, disagree very much). A 'feedback' statement is crafted for each level of response, for each statement (item in the scale). After a participant makes a selection of a response category in the questionnaire, the participant receives the message that was crafted to their response selection. There are two sources of error in such statements: the extent to which an item does not reflect the full scale and the extent to which the scale has error in measuring the underlying concept. Furthermore, our analyses showed that having multiple cut points on a continuum (in this case, four cut points across five response categories) increased the error. Given the extent to which an item does not validly capture the underlying concept, inappropriate information could be given to participants. Lower error per item may be expected within the Item Response Theory (IRT) framework, which selects items to be cut points along an underlying continuum[14]. Thus, IRT should guide the future development of measures in this area.

The results reported in this study are not meant to serve as criticism of current efforts to provide services to those in need but to illustrate the potential magnitude of misclassification associated with the measurement error inherent in brief assessments, and thereby provide a basis for improving the mechanism of distributing a programme's resources. Current brief dietary assessment methods such as those used to determine programme eligibility have not, and perhaps cannot, attain the low levels of measurement error necessary to achieve low rates of misclassification. Clearly, more research needs to be conducted to better understand how people report dietary data[15] and what could be done to improve these methods[16]. These issues have also been explicated in the context of epidemiological studies, showing the dramatic effect that even modest levels of error can have on detecting associations of risk[17]. Nevertheless, policy-makers and the public must decide how much misclassification error they are willing to accept when certifying people to receive (or not) government services. Perhaps dietary assessment for certification or even for education at the level of the individual should be discontinued and new possibilities explored.

A two-stage assessment procedure could be conducted. In this approach, the first stage of assessment may comprise a brief dietary assessment in which the observed cut point is set higher than the targeted true intake to maximise the percentage of truly eligible subjects who would be classified as eligible. This would be followed by a second stage of assessment, of only those applicants who met the first-stage criteria, that would include more lengthy dietary assessment such as the 24-hour recall or additional related measures. This second stage would then serve to eliminate subjects meeting the first criterion but who are truly not eligible. This approach has the advantage of providing more accurate classification without the increased cost in time and money associated with intensive assessment of every applicant. Similar approaches have been used in programmes for screening for sub-clinical disease in otherwise healthy populations, particularly when resources are limited[18].

## Acknowledgements

This project has been funded in part by federal funds from the US Department of Agriculture (USDA)/Agriculture Research Service (ARS) under Cooperative Agreement No. 58-6250-6001. This work is, in part, a publication of the USDA/ARS Children's Nutrition Research Center, Department of Pediatrics, Baylor College of Medicine and Texas Children's Hospital, Houston, Texas. The contents of this publication do not necessarily reflect the views or policies of the USDA, nor does mention of trade names, commercial products or organisations imply endorsement by the US Government. The authors would like to thank Dr Gladys Block and Dr Valerie Tarasuck for their comments on an earlier draft of this manuscript.

## References

1 Institute of Medicine. *WIC Nutrition Risk Criteria: A Scientific Assessment*. Washington, DC: National Academy Press, 1996.
2 Fox MK, Burstein N, Golay J, Price C. *WIC Nutrition Education Assessment Study, Final Report*. Cambridge, MA: Abt Associates, 1998.
3 Di Clemente CC, Marinhilli AS, Singh M, Bellino LE. The role of feedback in the process of health behavior change. *American Journal of Health Behavior* 2001; **25**: 217–27.
4 Gersovitz M, Madden JP, Smiciklas-Wright H. Validity of the 24-hour dietary recall and seven day record for group comparison. *Journal of the American Dietetic Association* 1978; **73**: 48–55.
5 Block G, Thompson FE, Hartman AM, Larkin FA, Guire KE. Comparison of two dietary questionnaires validated against multiple dietary records collected during a 1-year period. *Journal of the American Dietetic Association* 1992; **92**: 686–93.
6 Block G, Subar AF. Estimates of nutrient intake from a food frequency questionnaire: the 1987 National Health Interview

Survey. *Journal of the American Dietetic Association* 1992; **92**: 969–77.
7 Willett W. *Nutritional Epidemiology*. New York: Oxford University Press, 1990.
8 Day NE, McKeown N, Wong MY, Welch A, Bingham S. Epidemiological assessment of diet: a comparison of a 7-day diary with a food frequency questionnaire using urinary markers of nitrogen, potassium and sodium. *International Journal of Epidemiology* 2001; **30**: 309–17.
9 Kipnis V, Carroll RJ, Freidman LS, Li L. Implications of a new dietary measurement error model for estimation of relative risk: application to four contribution studies. *American Journal of Epidemiology* 1999; **150**: 642–51.
10 Scholl TO, Hediger ML, Schall JI, Khoo CS, Fischer RL. Dietary and serum folate: their influence on the outcome of pregnancy. *American Journal of Clinical Nutrition* 1996; **63**: 520–5.
11 Troiano RP, Flegal KM, Kuczmarski RJ, Campbell SM, Johnson CL. Overweight prevalence and trends for children and adolescents: The National Health and Nutrition Examination Surveys, 1963–1991. *Archives of Pediatrics and Adolescent Medicine* 1995; **149**: 1085–91.
12 Kipnis V, Midthune D, Freedman LS, Bingham S, Schatzkin A, Subar A, *et al*. Empirical evidence of correlated biases in dietary assessment instruments and its implications. *American Journal of Epidemiology* 2001; **152**: 394–403.
13 Randall B, Bartlett S, Kennedy S. *Study of WIC Participant and Program Characteristics, 1996*. FNS 53-3198-3-026. Washington, DC: US Department of Agriculture, Food and Nutrition Service, Office of Analysis and Evaluation, 1998.
14 Humbleton RK, Swaminathan H, Rogers HJ. *Fundamentals of Item Response Theory*. Newbury Park, CA: Sage, 1991.
15 Baxter SD, Thompson WO, Davis HC, Johnson MH. Impact of gender, ethnicity, meal component, and time interval between eating and reporting on accuracy of fourth-graders' self-reports of school lunch. *Journal of the American Dietetic Association* 1997; **97**: 1293–8.
16 Munoz KA, Krebs-Smith SM, Ballard-Barbash R, Cleveland LE. Food intakes of US children and adolescents compared with recommendations. *Pediatrics* 1997; **100**: 323–9.
17 Walker AM, Blettner M. Comparing imperfect measures of exposure. *American Journal of Epidemiology* 1985; **121**: 783–90.
18 Kuhn DL, Risi L, Richart RM, Pollack A, Kostecki F, Wright TC. Two-stage cervical cancer screening: an alternative for resource-poor settings. *American Journal of Obstetrics and Gynecology* 2000; **183**: 383–8.