

promoting access to White Rose research papers



Universities of Leeds, Sheffield and York
<http://eprints.whiterose.ac.uk/>

This is an author produced version of a paper published in **American Journal of Epidemiology**.

White Rose Research Online URL for this paper:

<http://eprints.whiterose.ac.uk/75638/>

Published paper:

Greenwood, DC, Ransley, J, Gilthorpe, MS and Cade, JE (2006) *Use of Itemized Till Receipts to Adjust for Correlated Dietary Measurement Error*. American Journal of Epidemiology, 164 (10). 1012 - 1018.

<http://dx.doi.org/10.1093/aje/kwj308>

Use of itemised till receipts to adjust for correlated dietary measurement error

Greenwood DC¹, Ransley JK², Gilthorpe MS¹, Cade JE².

Abbreviations: CI, confidence interval

¹ Biostatistics Unit, Centre for Epidemiology & Biostatistics, University of Leeds, 30-32 Hyde Terrace, Leeds. LS2 9LN. UK

² Nutrition Epidemiology Group, Centre for Epidemiology & Biostatistics, University of Leeds, 30-32 Hyde Terrace, Leeds. LS2 9LN. UK

Address for correspondence: Biostatistics Unit, Centre for Epidemiology & Biostatistics, University of Leeds, 30-32 Hyde Terrace, Leeds. LS2 9LN. UK. (e-mail: d.c.greenwood@leeds.ac.uk)

Running head: Till receipts and dietary measurement error

Abstract word count: 142

Text word count: 3997

ABSTRACT

Recent studies suggest that measurement error in food frequency questionnaires includes a person-specific component correlated with that of other self-reported dietary assessments. Use of biomarkers has been recommended to adequately calibrate dietary assessment tools for unbiased estimation of associations between diet and disease. Biomarkers of intake are often only collected on small sub-samples because they can be expensive and inconvenient for participants. The authors propose a novel approach using household itemized till receipts to calibrate dietary assessment. Till receipts are not self-recorded and not subject to a person-specific bias but need to be supported by self-completed diaries for food eaten away from home. They may also prove cheaper to collect on larger samples. The authors discuss the many methodological challenges of using household level data, and discuss how till receipts might be used in practice, with or without the use of biomarkers.

Biological markers; diet; diet surveys; epidemiologic methods; nutrition surveys; questionnaires

ACKNOWLEDGEMENTS

Early work on using itemized till receipts as dietary measures was funded by the UK Medical Research Council and UK Department of Health.

Common methods of adjustment for measurement error assume that measurement errors in the reference instrument are independent of those in the error-prone tool being calibrated (1). Recent research indicates that use of multiple 24-hour recalls or food diaries covering a number of days are not adequate reference instruments to calibrate food frequency questionnaires (FFQs) because all dietary assessment tools based on self-report are subject to measurement error that is correlated with that of other self-report tools (2, 3). Studies using biomarkers of dietary intake suggest that individuals may differ systematically in the accuracy and precision of their reporting (4-8). Several measurement error models have been suggested that take advantage of biomarkers (4, 9-14). In particular a new measurement error model allowing for person-specific bias in self-report instruments has been proposed, using information from FFQ, 24-hour recalls or food diaries, and biomarkers of intake (15) and minimal requirements specified for validation studies (8). Measurement errors in biomarkers can reasonably be assumed to be independent of those of the self-reported tools, because they are obtained independently. Use of biomarkers suggests that bias caused by measurement error may be twice as strong as that estimated using self-report tools alone (6, 7).

Despite the important advantages of using biomarkers, there are some problems with their use for calibrating tools assessing dietary intake. Firstly, whilst there are many good biomarkers of exposure at the cellular level, there are only a handful of biomarkers that adequately reflect intake. Biomarkers that predict, in an unbiased manner, the true intake of a particular dietary component include doubly labelled water for total energy intake, urinary nitrogen for protein intake, urinary potassium and sodium for those mineral intakes (5, 16, 17). Most other biomarkers do not give a

clear, strong, relation with intake, unrelated to individual characteristics, and with errors unrelated to the true intake, so do not meet the requirements for calibration (7, 10, 14, 16-23). If any other nutrient is required as a predictor in a regression model, either as the main exposure, confounder or effect modifier, then measurement error from this source cannot be eliminated by the use of a biomarker because no adequate calibration is available for other nutrients.

Even valid biomarkers are subject to a large amount of random variation compared to the dietary intake of relevance to the outcome. This will not so much be laboratory error, but more likely result from day-to-day variation in diet. In epidemiology it is likely that a long term measure of diet is required, or intake earlier in life, whereas biomarkers give only a small snapshot of current intake on a particular day (16). This leads to estimates corrected for measurement error, but with much wider confidence intervals than the uncorrected estimates. Furthermore, these measures are expensive and invasive to collect (16). It is not feasible to collect them on any more than a small subgroup of a cohort study.

In this paper we therefore propose a new hierarchical model for dietary measurement error based on relatively objective household till receipts, reducing the problem of correlated person-specific biases. Household till receipts share the property with biomarkers of not being self-report measures, thereby avoiding correlated person-specific biases. We discuss the advantages of the method over the use of biomarkers alone, and illustrate the application of the method using simulated data. We also outline some challenges with the application of this method and discuss possible solutions.

METHODS

Deriving household diet from till receipts

Itemized till receipts provide a prospective record of food products purchased by a household. They contain sufficient information to identify the exact products purchased, from which the nutrient content can be derived in a similar manner to food diaries or detailed 24-hour recalls, based upon standard food databases (24, 25). Pet foods and non-food items are excluded. A record of visitors attending meals, meals eaten away from home, and food purchased from shops not providing itemized till receipts may be necessary. Detailed methods are presented elsewhere (26, 27).

Several factors make it difficult to derive information from till receipts at an individual level. First, different household members will consume different proportions of the household diet, for example adults will eat more than children. Second, a proportion of an individual's diet may be consumed outside of the household, or without an itemized till receipt. Third, household visitors may consume a proportion of the food purchased. Fourth, bulk purchases will add a potentially large component of random error to the measurement, for example for food purchased for the freezer, cooking oil or alcoholic beverages.

Disease model

In outlining the method, we follow the notation of Kipnis et al.(15) where possible.

Firstly, consider the disease model: $R(D | T) = \alpha_0 + \alpha_1 T$ (1)

where $R(D | T)$ is the risk of disease outcome D on an appropriate scale such as the logit, conditional on T , the true dietary intake of relevance to developing the disease, such as the true long-term intake, or intake during an “at risk” period; α_0 is a constant and α_1 is the parameter of interest representing the strength of association between true dietary intake and the disease.

FFQ model and reference instrument model

We consider household h , individual i , period or season j , and replicate k . True intake is not known, but we have diet measured imperfectly by a FFQ Q_{hij} , a reference instrument F_{hij} such as a 24-hour recall or food diary, and a biomarker M_{hij} . We model the FFQ and the dietary reference instrument in a similar manner to Kipnis *et al.* (15) and Spiegelman *et al.* (8).

$$Q_{hij} = \mu_{Qj} + \beta_{Q0} + \beta_{Q1}T_{hi} + r_{hi} + \varepsilon_{hij} \quad (2)$$

$$F_{hij} = \mu_{Fj} + \beta_{F0} + \beta_{F1}T_{hi} + s_{hi} + u_{hij} \quad (3)$$

where μ_{Qj} and μ_{Fj} represent a possible drift over the time period between measures, or a seasonal effect (28, 29) in order to improve model fit; β_{Q0} , β_{Q1} , β_{F0} and β_{F1} are biases where β_{Q0} and β_{F0} are additive components associated with the instruments used, and β_{Q1} and β_{F1} are multiplicative components; r_{hi} and s_{hi} model the person-specific bias for each tool. We allow these person-specific biases to be correlated, with correlation $\rho(r,s) \neq 0$, because the same mechanisms may be influencing both r_{hi} and s_{hi} . We assume within-person errors ε_{ij} and u_{ij} are independent of each other and follow normal distributions with zero mean and variances σ_ε^2 and σ_u^2 respectively.

The error terms ε_{hij} and u_{hij} include any deviation between short-term and long-term intake. It would be possible to allow for correlation between ε_{hij} and u_{hij} within the same season, but this has previously been demonstrated to be negligible (6).

Till receipt model

We propose modeling till receipt L_{hj} for household h and season j as:

$$L_{hj} = \frac{1}{1-c_h} \sum_i (T_{hij} + \mu_{Lj} + z_{hi}) + \xi_{hj} \quad (4)$$

and

$$\frac{T_{hi}}{\sum_i T_{hi}} = \pi_{hi} \quad (5)$$

where $1-c_h$ represents the proportion of purchased food that is eventually eaten by the household, and c_h represents the proportion of household food wastage; μ_{Lj} is a possible seasonal effect; ξ_{hj} is the household-level error term, independent of the other error terms, following a normal distribution with zero mean and variance σ_ξ^2 ; and π_{hi} represents the proportion of the till receipt attributable to individual i in household h . In keeping with the analogous biomarker model proposed by Kipnis *et al.* we assume that the person-specific bias z_{hi} is negligible, because of the objective prospective nature of the data collection, and can therefore be assumed to be zero.

Use of biomarkers

It is necessary to derive the proportions π_{hi} and c_h either from other data or by making assumptions regarding their distribution. Estimates might be obtained from large national surveys such as the NDNS (Gregory et al. (30)), providing these have been adequately validated using non self-report measures. Alternatively, estimates might be derived from within the same study by use of a biomarker M_{hijk} with k repeat measurements within season j .

$$M_{hijk} = \mu_{Mj} + T_{hi} + w_{hi} + v_{hijk} \quad (6)$$

with proportions π_{hi} and c_h now estimated from equations (4) and (5).

In keeping with Kipnis *et al.* we assume that the person-specific bias w_{hi} is negligible, and can be assumed to be zero, and that the within-person error v_{hijk} is random and independently distributed. Any of these models could easily be extended to allow for heterogeneity in the study population due to age, sex or body mass if necessary (31).

Estimation with biomarkers on just one household member

Ideally, the method outlined here would use biomarkers collected from the whole household so that estimates of π_{hi} (proportion of household intake consumed by an individual) and c_h (proportion of wastage) could be derived directly from biomarkers. Most epidemiological studies, however, do not include all members of a household. If wastage could be derived from prior knowledge, previous surveys, external data, or trusted to self-report, then (ignoring any seasonal effect) π_{hi} can be estimated from

$$\frac{M_{hi}}{(1-c_h) \sum_i T_{hi}} = \pi_{hi} \quad (7)$$

Alternatively, from a previous study it may be possible to model π_{hi} based on, say, age and sex, and then use this to estimate the proportion in the current study.

Estimation in the absence of biomarkers

Initially there would appear little advantage in collecting itemized till receipts if biomarkers were required to derive π_{hi} and c_h . However, validation against biomarkers need only be performed once for a given population and thereafter π_{hi} and c_h may be considered known. Another consideration is that for most food and nutrient intakes, no appropriate biomarkers of intake exist. However, if we can assume that the proportion of food purchased by each individual within a household is consistent across different exposures, such that π_{hi} for one exposure and π_{hi}^* for a second exposure are equal, then only one biomarker would be required to estimate π_{hi} for all exposures of interest. Further work is required with real data to demonstrate whether this strong assumption is better than having no objective standard with which to calibrate self-report measures and leaving the associated problem of correlated measurement error unresolved.

An alternative source for estimating π_{hi} for foods and nutrients without an appropriate biomarker would be to assume that, whilst the absolute intake derived from a self-report measure is subject to a person-specific bias, the proportion π_{hi} is not. With this assumption π_{hi} could be derived from the reference instrument such as the food diaries or 24-hour recalls

$$\frac{F_{hi}}{\sum_i F_{hi}} = \pi_{hi} \tag{8}$$

or even from the FFQs

$$\frac{Q_{hi}}{\sum_i Q_{hi}} = \pi_{hi} \quad (9)$$

If it were reasonable to assume that person-specific bias associated with the reference instrument, s_{hi} , or the FFQ, r_{hi} , could be replaced by household-specific bias s_h or r_h then the above equations would be valid. This is the same as saying that characteristics shared by the household influence the self-reported diet, but conditional on this, not at the individual level. In reality household-level characteristics are likely to form part, but not all, of the person-specific bias, but further work with real data is required to show whether this is still better than not correcting for any of the person-specific bias.

Model fitting

The method of maximum likelihood can be used to estimate parameters, or Markov chain Monte Carlo (MCMC) methods within a Bayesian framework (32).

SIMULATIONS

To our knowledge, no dataset exists with till receipts and biomarkers collected on the same individuals. We therefore illustrate our model on a series of simulations based on investigating the association between protein intake and breast cancer incidence.

Data were sampled from distributions with similar means and variances to those reported by Kipnis *et al.* (7, 33) adapted to incorporate household till receipt measures. Household structure was generated to be broadly similar to a previous study

(26, 27), with 600 individuals (200 adult men, 200 adult women, and 200 children) allocated at random to one of 200 households. For the purposes of this simple illustration we assumed the mean intakes of women and children to be 80 percent and 50 percent of a man's respectively, so that $\beta_{h,male}=1.0$, $\beta_{h,female}=0.8$ and $\beta_{h,child}=0.5$ for all h , and we also assumed that 10 percent of food purchased was not eaten, so that $c_h = 0.1$ for all h . All measurements are log-transformed to allow additive and homoscedastic measurement errors for biomarkers (7, 33). We assumed the mean (standard deviation) log-transformed intake for adult males was 4.5 (0.2) to give a geometric mean protein intake of 90 g/day. We allow for a small drift in recorded intakes of 0.06 between two FFQs and 0.02 between two 24-hour recalls, so that: $\mu_{Qj} = 0.06$, $\mu_{Fj} = 0.02$, whilst $\mu_{Lj} = \mu_{Mj} = 0$. Additive and multiplicative components of reporting biases in the tools were set to reflect an underestimation of the food frequency questionnaire and 24-hour recalls: $\beta_{Q0} = 1.25$, $\beta_{Q1} = 0.65$, $\beta_{F0} = 1.4$, and $\beta_{F1} = 0.65$, though the statistical methods would apply just as well if the biases acted in different directions. Person-specific biases were also included: $\sigma_r^2 = 0.35$, $\sigma_s^2 = 0.18$, $\rho(r,s) = 0.3$. Error variances were $\sigma_\varepsilon^2 = 0.21$, $\sigma_u^2 = 0.33$, $\sigma_\xi^2 = 0.33$, and $\sigma_v^2 = 0.11$ (the latter based on the estimated error variance for 28 24-hour recalls). In the disease model the intercept was $\alpha_0 = -3$, and the slope $\alpha_1 = 0.7$, chosen to give a realistic odds ratio of approximately 2.0 for comparison of the highest to the lowest quartile of intake.

Simulated data were generated using Stata 9.1 (34). The models were fitted within the Bayesian framework using WinBugs 1.4.1 (35) called from within Stata. All stochastic parameters were given proper but minimally informative prior distributions. Convergence appeared to be achieved after a 20,000 update burn-in, for each of two

chains with dispersed initial values. This was followed by a further 10,000 updates for each chain. Adequate mixing and convergence was confirmed by assessment of trace plots and Brooks-Gelman-Rubin statistics (36), with the Monte Carlo error for each parameter of interest less than 5 percent of the sample standard deviation. To allow for random sampling error in simulating the data, this process was replicated 100 times, with the mean and empirical standard deviation of the estimates compared to true values.

Eight measurement error models were compared, designed to reflect different potential analytical strategies:

- (i) To demonstrate the bias introduced by measurement error, we consider a naïve analysis ignoring measurement error in a single FFQ. This reflects common practice in many studies. A simple logistic regression model for the association between a single measure of protein intake and breast cancer incidence is used.
- (ii) a logistic regression with a simple adjustment for measurement error using a second measure of protein intake derived from a replicate FFQ, with no allowance for correlated person-specific biases.
- (iii) a logistic regression with a simple adjustment for measurement error using a more accurate measure of protein intake derived from a 24-hour recall, but again with no allowance for correlated person-specific biases.
- (iv) a logistic regression model using two FFQs, two 24-hour recalls (or food diaries), as in equations (2) and (3), with two measures of urinary nitrogen as a biomarker for protein intake. The model allows for correlated person-specific bias, as in Kipnis *et al.* (6), by calibrating against the objective biomarkers that have negligible person-specific bias.

(v) use of repeat FFQs, 24-hour recalls (or diaries) and till receipts, with allowance for correlated person-specific bias, assuming the proportions π_{hi} and c_h are perfectly known. This represents a model that might be used if no biomarkers were available in the study or for a particular nutrient.

(vi) a logistic regression model using two FFQs, two 24-hour recalls (or food diaries), as in equations (2) and (3), with two 28-day collections of itemized till receipts. The model allows for correlated person-specific bias, as in (4), by calibrating against the protein intake derived from till receipts with proportions π_{hi} and c_h derived from two measures of urinary nitrogen as a biomarker for protein intake included in the same model. This represents a model that could be used if some biomarker measures are available.

(vii) to explore the sensitivity of model (v) to incomplete recording of intake by till receipts, based on previous work (26, 27) we assumed that 12% of dietary intake was not captured by itemized till receipts, but recorded by 28-day shopping diaries subject to the same person-specific bias as 24-hour recalls.

(viii) to explore the sensitivity of model (vi) to incomplete till receipts we assumed that 12% of dietary intake was not captured by itemized till receipts, but recorded by 28-day shopping diaries subject to the same person-specific bias as 24-hour recalls.

Ignoring measurement error more than halved the slope (log-odds ratio) from 0.70 to 0.28, reducing an odds ratio of 2.0 to 1.3 (95 percent confidence interval (CI): 0.9, 1.9) (Table 1). Using a repeat FFQ also led to the effects of measurement error being underestimated, with the estimated coefficient still half its true value. Using a 24-hour recall was substantially better than using a repeat FFQ for adjusting for measurement error. Using a biomarker alongside the FFQ and 24-hour recall leads to improved

estimates (estimates within one standard error of the true values), but with slightly larger standard errors. Using till receipts alongside the 24-hour recall and FFQ, with a biomarker to estimate the proportion of food purchased consumed by each individual, π_{hi} , and the proportion of food purchased that is consumed by each household, c_h , also gave good estimates within one standard error of the true values of the parameters within our simulated data. Using till receipts and assuming the proportions π_{hi} and c_h were known gave similar estimates in our simulation, with slightly smaller standard errors, without the need for biomarkers.

Sensitivity analysis of the robustness of the models to realistic assumptions regarding incomplete collection of itemized till receipts suggested that use of till receipts was still better than diaries for calibration.

DISCUSSION

The feasibility of collecting itemized till receipts from households has been demonstrated previously (26, 27, 37). Till receipt collections are common in household budget surveys and market research. These receipts provide a prospectively recorded list of food products purchased and contain sufficient information to identify the exact products purchased, from which the nutrient content can be derived in a similar manner to food diaries or detailed 24-hour recalls. Methodologically, there are parallels with occupational epidemiology where an accurate measure of an occupational exposure may be available at a group level, such as a factory or job role, with less accurate information available for individuals (38-40).

Like self-report measures, use of till receipts to measure intake is subject to completeness of food tables (41). This may be a limitation for their use with some nutrients. However, the most important methodological issue is the completeness of till receipt collections.

A proportion of food consumed will have been purchased without an associated itemized till receipt. Although 85% of UK grocery shopping in 2000 was purchased in supermarkets (42), and many of the remaining smaller shops use itemized till receipts too, in the same year 9% of the weekly spend was in restaurants and cafés (43), with only some providing itemized bills. Consideration must therefore be given to meals eaten out of the home, food purchased from shops not providing itemized receipts, such as staff canteens, as well as guests eating with the household. It will probably be necessary to ask individuals to record a diary of meals eaten away from the home to support the information provided by the till receipts. This would be analyzed in the same way as a food diary to derive estimated nutrients based on standard portion sizes, lacking the precision of a weighed intake. Use of any additional self-reported record of intake such as this reduces the objectivity of the methodology and introduces an unwanted element of person-specific bias, albeit less than with a wholly self-reported measure. Because of this it probably not appropriate to consider use of itemized till receipts as a totally objective measure, but more objective than use of food diaries or 24-hour recalls alone. Sensitivity analysis of the robustness of the models to a realistic proportion of food consumed without an itemized receipt suggested that use of receipts still gave substantially better estimates of the diet-disease association than

using self-report measures in terms of both bias and precision, although the proportion of food wasted was underestimated.

We have discussed a number of strong assumptions that would allow the methods to be applied to the situation where no adequate biomarkers are available. For example, assuming that the same proportions π_{hi} hold across a range of different exposures. This implies that different members of the household eat meals of identical content and only the size of the meal varies. This may be an inappropriate assumption if, say, children do not eat their vegetables, or the men eat larger portions of meat than the rest of the household, even allowing for different overall meal sizes. The need to make these assumptions weakens the usefulness of the method. Further research is required to tell if records of additional meals and strong assumptions regarding proportion attributable to individuals in a household render the method no better than calibration against a purely self-report measure such as a food diary or 24-hour recall.

Bulk purchases for storage such as multi-packs, food for home-freezing, large containers of cooking oil, alcoholic beverages, etc. are characteristic of modern shopping habits, with over half of UK consumers bulk-buying (44). Similarly, households may store considerable quantities of food in a pantry, cupboard or freezer for later consumption, to the extent of requiring substantial storage space (44). Such purchasing and consumption patterns do not lessen the objectivity of the tool, and in the long run will balance out. However, they do add a potentially large component of random error to intake estimated from itemized till receipts. Further work is needed to explore alternative strategies to reduce the influence of stored foods. These could include pantry inventories at the start and end of a period of till receipt collection.

For FFQs, the amount of measurement error will depend on characteristics of the individual tool used, such as the number of items recorded, the assessment of portion size, whether frequency was categorized, and such like. Another advantage of using itemized till receipts is that the amount of measurement error in them does not depend on these characteristics. Therefore, if measurement error variances σ_{ξ}^2 were derived for receipts covering a particular time period, then these could be considered more transportable than equivalent variances for FFQs where the variance would depend more closely on the design of the particular FFQ.

In practice biomarkers and reference instruments are only collected from a sub-sample of the study for reasons of costs. Use of biomarkers can also be invasive and require substantial staff time collecting and analyzing samples, whilst instruments such as weighed food diaries and 24-hour recalls require substantial nutritionist coding time. Coding of till receipts also requires nutritionists' time, though there is the potential for this to be more automated if access to supermarket databases is available or if receipts are scanned and optical character recognition software used. It may be feasible to collect till receipts on a larger sub-sample than possible with the biomarkers, increasing precision of the final estimate. Alternatively, till receipts might provide an appropriate instrumental variable to augment a single biomarker measure, allowing the reliability ratio to be estimated (8).

In summary, our suggested method may require support by self-recorded diaries of meals not covered by the receipts, that reduce the objectivity of the method. Using till receipts may require strong assumptions to derive estimated intake for individuals

from the household-level data that weaken the usefulness of the method. Despite these substantial reservations, where adequate biomarkers do not exist, or are prohibitively expensive, we propose that using itemized till receipts provides a possible method for assessing diet that is less prone to correlated person-specific biases associated with self-report instruments. This allows for more complete adjustment for the effects of measurement error in estimating associations between diet and disease, with potentially tighter confidence intervals than those associated with the use of biomarkers prone to large random variation in small validation samples.

Table 1. Logistic regression coefficients, $\hat{\alpha}_0$ and $\hat{\alpha}_1$, estimated reliability ratio for the FFQ, $\hat{\lambda}_{FFQ}$, and correlation of correlation person-specific biases, $\hat{\rho}(r, s)$, for different measurement error models with empirical standard deviations of the estimates in parentheses.

		$\hat{\alpha}_0$	$\hat{\alpha}_1$	$\hat{\lambda}_{FFQ}$	$\hat{\rho}(r, s)$	\hat{c}_h	$\hat{\beta}_{male}$	$\hat{\beta}_{female}$	$\hat{\beta}_{child}$
(i)	Logistic regression ignoring measurement error	-1.17 (0.76)	0.28 (0.19)	1	0	-	-	-	-
(ii)	Repeat FFQ	-1.39 (0.95)	0.34 (0.24)	0.80 (.02)	0	-	-	-	-
(iii)	FFQ and recall	-2.56 (1.63)	0.60 (0.40)	0.32 (0.04)	0	-	-	-	-
(iv)	Biomarker, FFQ and recall	-3.19 (1.22)	0.74 (0.29)	0.37 (0.03)	0.23 (0.07)	-	-	-	-
(v)	Till receipts, FFQ and recall with π_{hi} and c_h known (assuming till receipts capture all dietary intake)	-3.10 (1.18)	0.69 (0.26)	0.29 (0.02)	0.42 (0.07)	0.1	1	0.8	0.5
(vi)	Biomarker, till receipts, FFQ and recall (assuming till receipts capture all dietary intake)	-3.16 (1.10)	0.74 (0.26)	0.36 (0.03)	0.31 (0.08)	0.10 (0.01)	1	0.84 (0.02)	0.58 (0.02)
(vii)	Till receipts, FFQ and recall with π_{hi} and c_h known (assuming diaries used to supplement till receipts for food consumed without receipts)	-2.89 (1.08)	0.71 (0.27)	0.29 (0.02)	0.40 (0.06)	0.1	1	0.8	0.5
(viii)	Biomarker, till receipts, FFQ and recall (assuming diaries used to supplement till receipts for food consumed without receipts)	-3.02 (1.19)	0.71 (0.28)	0.36 (0.03)	0.30 (0.08)	0.02 (0.01)	1	0.83 (0.02)	0.58 (0.02)

True values used in simulations are $\alpha_0=-3$, $\alpha_1=0.7$, $\lambda_{FFQ}=0.3$, $\rho(r,s)=0.3$, $c_h=0.1$, $\beta_{male}=1$, $\beta_{female}=0.8$, and $\beta_{child}=0.5$.

REFERENCES

1. Spiegelman D, Schneeweiss S, McDermott A. Measurement error correction for logistic regression models with an "alloyed gold standard". *American Journal of Epidemiology* 1997;145:184-96.
2. Hebert JR, Clemow L, Pbert L et al. Social desirability bias in dietary self-report may compromise the validity of dietary intake measures. *Int J Epidemiol* 1995;24:389-98.
3. Michels KB, Bingham SA, Luben R et al. The Effect of Correlated Measurement Error in Multivariate Models of Diet. *Am J Epidemiol* 2004;160:59-67.
4. Kaaks R, Riboli E, Esteve J et al. Estimating the accuracy of dietary questionnaire assessments - validation in terms of structural equation models. *Stat Med* 1994;13:127-42.
5. Day NE, McKeown N, Wong MY et al. Epidemiological assessment of diet: a comparison of a 7-day diary with a food frequency questionnaire using urinary markers of nitrogen, potassium and sodium. *Int J Epidemiol* 2001;30:309-17.
6. Kipnis V, Midthune D, Freedman LS et al. Empirical evidence of correlated biases in dietary assessment instruments and its implications. *Am J Epidemiol* 2001;153:394-403.
7. Kipnis V, Subar AF, Midthune D et al. Structure of Dietary Measurement Error: Results of the OPEN Biomarker Study. *Am J Epidemiol* 2003;158:14-21.

8. Spiegelman D, Zhao B, Kim J. Correlated errors in biased surrogates: study designs and methods for measurement error correction. *Stat Med* 2005;24:1657-82.
9. Fraser GE. A search for truth in dietary epidemiology. *Am J Clin Nutr* 2003;78:521S-525.
10. Bingham SA. Biomarkers in nutritional epidemiology. *Public Health Nutr* 2002;5:821-7.
11. Bingham SA. Urine Nitrogen as a Biomarker for the Validation of Dietary Protein Intake. *J Nutr* 2003;133:921S.
12. Fraser GE, Stram DO. Regression calibration in studies with correlated variables measured with error. *Am J Epidemiol* 2001;154:836-44.
13. Spiegelman D, Rosner B, Logan R. Estimation and inference for logistic regression with covariate misclassification and measurement error in main study/validation study designs. *J Am Stat Assoc* 2000;95:51-61.
14. Marshall JR. Methodologic and Statistical Considerations Regarding Use of Biomarkers of Nutritional Exposure in Epidemiology. *J Nutr* 2003;133:881S.
15. Kipnis V, Carroll RJ, Freedman LS et al. Implications of a new dietary measurement error model for estimation of relative risk: Application to four calibration studies. *Am J Epidemiol* 1999;150:642-51.
16. Bates CJ et al. Biochemical markers of nutrient intake. In: Margetts BM, Nelson M, eds. *Design concepts in nutritional epidemiology*. Oxford: Oxford University Press, 1997:170-240.

17. Prentice RL. Dietary assessment and the reliability of nutritional epidemiology reports. *Lancet* 2003;**362**:182-3.
18. Kaaks R. Biochemical markers as additional measurements in studies of the accuracy of dietary questionnaire measurements: conceptual issues. *Am J Clin Nutr* 1997;65 (Suppl):1232S-9S.
19. Kipnis V, Kipnis V, Midthune D et al. Bias in dietary-report instruments and its implications for nutritional epidemiology. *Public Health Nutr* 2002;5:915-23.
20. Taren D, Dwyer J, Freedman L et al. Dietary assessment methods: where do we go from here? *Public Health Nutr* 2002;5:1001-3.
21. Pollard J, Wild CP, White KLM et al. Comparison of plasma biomarkers with dietary assessment methods for fruit and vegetable intake. *Eur J Clin Nutr* 2003;57:988-98.
22. Mayne ST. Antioxidant Nutrients and Chronic Disease: Use of Biomarkers of Exposure and Oxidative Stress Status in Epidemiologic Research. *J Nutr* 2003;133:933S.
23. Nelson M. The validation of dietary assessment. In: Margetts BM, Nelson M, eds. *Design concepts in nutritional epidemiology*. Oxford: Oxford University Press, 1997:241-72.
24. Royal Society of Chemistry and Ministry of Agriculture FaF. McCance and Widdowson's the Composition of Foods. Cambridge: Royal Society of Chemistry, 1991.

25. U.S.Department of Agriculture ARS. USDA National Nutrient Database for Standard Reference, Release 18. Nutrient Data Laboratory Home Page. 2005.
26. Ransley JK, Donnelly JK, Khara TN et al. The use of supermarket till receipts to determine the fat and energy intake in a UK population. *Public Health Nutr* 2001;4:1279-86.
27. Ransley JK, Donnelly JK, Botham H et al. Use of supermarket receipts to estimate energy and fat content of food purchased by lean and overweight families. *Appetite* 2003;41:141-8.
28. Landin R, Freedman LS, Carroll RJ. Adjusting for time trends when estimating the relationship between dietary-intake obtained from a food frequency questionnaire and true average intake. *Biometrics* 1995;51:169-81.
29. Carroll RJ, Ruppert D, Stefanski LA. Measurement error in nonlinear models. London: Chapman & Hall, 1995.
30. Gregory J et al. The dietary and nutritional survey of British adults. London: H. M. Stationery Office, 1998.
31. Willett W. Commentary: Dietary diaries versus food frequency questionnaires - a case of undigestible data. *Int J Epidemiol* 2001;30:317-9.
32. Gilks WR, Richardson S, Spiegelhalter DJ. Markov chain monte carlo in practice. London: Chapman & Hall, 1996.
33. Subar AF, Kipnis V, Troiano RP et al. Using Intake Biomarkers to Evaluate the Extent of Dietary Misreporting in a Large Sample of Adults: The OPEN Study. *Am J Epidemiol* 2003;158:1-13.

34. StataCorp. Stata statistical software: Release 9.1. College Station, TX: Stata Corporation, 2005.
35. Spiegelhalter DJ et al. BUGS: Bayesian inference Using Gibbs Sampling, Version 0.50. Cambridge: MRC Biostatistics Unit, 1995.
36. Brooks SP, Gelman A. Alternative methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics* 1998;7:434-55.
37. DeWalt KM, D'Angelo S, McFadden M et al. The use of itemized register tapes for analysis of household food acquisition patterns prompted by children. *J Am Diet Assoc* 1990;90:559-62.
38. Richardson S. Measurement error. In: Gilks WR, Richardson S, Spiegelhalter DJ, eds. *Markov chain monte carlo in practice*. London: Chapman & Hall, 1996:401-17.
39. Richardson S, Best N. Bayesian hierarchical models in ecological studies of health- environment effects. *Environmetrics* 2003;14:129-47.
40. Gilks WR, Richardson S. Analysis of disease risks using ancillary risk factors, with application to job-exposure matrices. *Stat Med* 1992;11:1443-63.
41. Cowin I, Emmett P. The effect of missing data in the supplements to McCance and Widdowson's food tables on calculated nutrient intakes. *Eur J Clin Nutr* 1999;53:891-4.

42. Competition Commission. Supermarkets: a Report on the Supply of Groceries from Multiple Stores in the United Kingdom. 2000. London, The Stationery Office.
43. Office of National Statistics. Family Spending. A report on the 1999-2000 Family Expenditure Survey. 2000. London, Office of National Statistics.
44. Mintel. Bulk buying in food and drink. 2004. London, Mintel International Group Limited.