Use of Itemized Till Receipts to Adjust for Correlated Dietary Measurement Error

D. C. Greenwood1, J. K. Ransley2, M. S. Gilthorpe1, and J. E. Cade2

1 Biostatistics Unit, Centre for Epidemiology and Biostatistics, University of Leeds, Leeds, United Kingdom.
2 Nutrition Epidemiology Group, Centre for Epidemiology and Biostatistics, University of Leeds, Leeds, United Kingdom.

Received for publication August 9, 2005; accepted for publication April 13, 2006.

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. Data on biomarkers of intake are often collected only in small subsamples, because collection of biomarker data can be expensive and inconvenient for participants. In this paper, the authors propose a novel approach using itemized household grocery till receipts to calibrate dietary assessment. Till receipts are not self-recorded and the data obtained from them are not subject to person-specific bias, but the data need to be supported by self-completed diaries for foods eaten away from home. Till receipts may also prove cheaper to collect in 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

Abbreviation: FFQ, food frequency questionnaire.

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 for calibrating 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 that take advantage of biomarkers have been suggested (4, 9–14). In particular, a new measurement-error model has been proposed that uses biomarkers along with information from FFQs and 24-hour recalls or food diaries to correct for person-specific bias in self-report instruments (15). Minimum requirements for such validation studies have also been described (8). Measurement errors in biomarkers can reasonably be assumed to be independent of those of self-report tools, because these measures 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, while 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 labeled water for total energy intake, urinary nitrogen for protein intake, and urinary potassium and sodium for those mineral intakes (5, 16, 17). Most other biomarkers do not have a
clear, strong relation with intake that is unrelated to individual characteristics and errors unrelated to true intake, and 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 or as a confounder or effect modifier, measurement error from this source cannot be eliminated by the use of a biomarker, because no adequate calibration is available for other nutrients.

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

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

METHODS

Deriving household dietary data from till receipts

Itemized grocery till receipts provide a prospective record of the 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 manner similar to that of food diaries or detailed 24-hour recalls, based on standard food databases (24, 25). Pet foods and nonfood items are excluded. A supplemental record of visitors attending household 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 grocery 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 (e.g., 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 scale, conditional on $T$, the true dietary intake of relevance to disease risk, such as true long-term intake or intake during an “at risk” period; $\alpha_0$ is a constant; and $\alpha_1$ 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 dietary data measured imperfectly by means of an 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 manner similar to that of Kipnis et al. (15) and Spiegelman et al. (8).

$$Q_{hij} = \mu_{Qj} + \beta_{Qj} T_{hi} + r_{hi} + \varepsilon_{hij}$$  

(2)

$$F_{hij} = \mu_{Fj} + \beta_{Fj} T_{hi} + s_{hi} + u_{hij},$$  

(3)

where $\mu_{Qj}$ and $\mu_{Fj}$ represent a possible drift over the time period between measures or a seasonal effect (28, 29), in order to improve model fit; $\beta_{Qj}$, $\beta_{Fj}$, and $\beta_{Tj}$ are biases, where $\beta_{Qj}$ and $\beta_{Fj}$ are additive components associated with the instruments used and $\beta_{Tj}$ are multiplicative components; and $r_{hi}$ and $s_{hi}$ model the person-specific bias for each tool. We allow these person-specific biases to be correlated, with correlation $r(s, t) \neq 0$, because the same mechanisms may be influencing both $r_{hi}$ and $s_{hi}$. We assume that within-person errors $\varepsilon_{hij}$ and $u_{hij}$ are independent of each other and follow normal distributions with means of zero and variances $\sigma^2_{\varepsilon}$ and $\sigma^2_u$, respectively.

The error terms $\varepsilon_{hij}$ and $u_{hij}$ include any deviation between short-term intake and long-term intake. It would be possible to allow for correlation between $\varepsilon_{hij}$ 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 $L_{hij}$ for household $h$ and season $j$ as

$$L_{hij} = \frac{1}{1-c_h} \sum_{i} \left( T_{hi} + \mu_{Lj} + \varepsilon_{hij} \right) + \tilde{\varepsilon}_{hij}$$  

(4)

and

$$\frac{T_{hi}}{\sum_i T_{hi}} = \pi_{hi},$$  

(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; $\mu_{Lj}$ is a possible
seasonal effect; $\hat{\zeta}_{hi}$ is the household-level error term, independent of the other error terms, following a normal distribution with mean zero and variance $\sigma_{\zeta}^2$, and $\pi_{hi}$ 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. (6), 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 $\pi_{hi}$ 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 Dietary and Nutritional Survey of British Adults (30), provided that these have been adequately validated using non-self-report measures. Alternatively, estimates might be derived from within the same study through the use of a biomarker $M_{hijk}$ with $k$ repeat measurements within season $j$.

$$M_{hijk} = \mu_{Mj} + T_{hi} + w_{hi} + \nu_{hijk},$$

(6)

with proportions $\pi_{hi}$ and $c_h$ now being estimated from equations 4 and 5.

In keeping with Kipnis et al. (6), we assume that the person-specific bias $w_{hi}$ is negligible and can be assumed to be zero, and that the within-person error $\nu_{hijk}$ 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 biomarker data on just one household member

Ideally, the method outlined here would use biomarker data collected from the whole household so that estimates of $\pi_{hi}$ (proportion of household intake consumed by an individual) and $c_h$ (proportion of wastage) could be derived directly from biomarkers. Most epidemiologic studies, however, do not include all members of a household. If wastage could be derived from prior knowledge, previous surveys, or external data or trusted to self-report, then (ignoring any seasonal effect) $\pi_{hi}$ could 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 $\pi_{hi}$ based on, for example, 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 to be little advantage in collecting itemized till receipts if biomarkers were required to derive $\pi_{hi}$ and $c_h$. However, validation against biomarkers need only be performed once for a given population; thereafter, $\pi_{hi}$ 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 $\pi_{hi}$ for one exposure and $\pi_{ih}^{*}$ for a second exposure are equal, then only one biomarker would be required to estimate $\pi_{hi}$ for all exposures of interest. Further work with real data will be required in order 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 $\pi_{hi}$ for foods and nutrients without an appropriate biomarker would be to assume that, while the absolute intake derived from a self-report measure is subject to person-specific bias, the proportion $\pi_{hi}$ is not. With this assumption, $\pi_{hi}$ could be derived from the reference instrument, such as the food diaries or 24-hour recalls:

$$\frac{F_{hi}}{\sum_i F_{hi}} = \pi_{hi}, \quad (8)$$

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 $r_h$, the above equations would be valid. This is the same as saying that characteristics shared by the household influence the self-reported diet, but beyond this an individual’s characteristics have no additional influence on the amount of measurement error. 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 methods within a Bayesian framework can be used (32).

SIMULATIONS

To our knowledge, there is no existing data set with data from both till receipts and biomarkers collected in the same persons. Therefore, we illustrate our model in a series of simulations based on investigating the association between protein intake and breast cancer incidence.

Data were sampled from distributions with mean values and variances similar 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 that of a previous study (26, 27), with 600 persons (200 adult men, 200 adult women, and 200 children) being 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 $\beta_{h,\text{male}} = 1.0$, $\beta_{h,\text{female}} = 0.8$, and $\beta_{h,\text{child}} = 0.5$ for all $h$. We also assumed that 10 percent of the food purchased was not eaten, so that $c_h = 0.1$ for all $h$. All measurements were log-transformed to allow additive and homoscedastic measurement errors for biomarkers (7, 33). We assumed that the mean log-transformed intake for adult males was 4.5 (standard deviation, 0.2), which gives a geometric mean protein intake of 90 g/day. We allowed for a small drift in recorded intakes of 0.2%, which gives a geometric mean protein intake of 90 g/day. We allowed for a small drift in recorded intakes of 0.2% between two FFQs and 0.02 between two 24-hour recalls, so that $\mu_{Oj} = 0.06$ and $\mu_{Fj} = 0.02$, while $\mu_{Lj} = \mu_{Mj} = 0$. Additive and multiplicative components of reporting biases in the tools were set to reflect an underestimation of the intakes derived from the FFQ and 24-hour recalls: $\beta_{Q0} = 1.25$, $\beta_{Q1} = 0.65$, $\beta_{F0} = 1.4$, and $\beta_{F1} = 0.65$—although the statistical methods would apply just as well if the biases acted in different directions. Person-specific biases were also included: $\sigma^2_h = 0.35$, $\sigma^2_r = 0.18$, and $\rho(s,r)^2 = 0.3$. Error variances were $\sigma^2_e = 0.21$, $\sigma^2_u = 0.33$, $\sigma^2_u = 0.33$, and $\sigma^2_r = 0.11$ (the latter being based on the estimated error variance for 24 24-hour recalls). In the disease model, the intercept was $\alpha_0 = -3$ and the slope was $\gamma_0 = 0.7$, chosen to produce a realistic odds ratio of approximately 2.0 for comparison of the highest quartile of intake with the lowest.

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 being 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 being compared with true values.

We compared eight measurement-error models designed to reflect different potential analytical strategies:

1. 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.

2. 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.

3. A logistic regression with a simple adjustment for measurement error using a more accurate measure of protein intake derived from a 24-hour recall, again with no allowance for correlated person-specific biases.

4. A logistic regression model using two FFQs and 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 the paper by Kipnis et al. (6), by calibrating against the objective biomarkers that have negligible person-specific bias.

5. Use of repeat FFQs, 24-hour recalls (or diaries), and till receipts, with allowance for correlated person-specific bias, assuming that the proportions $\pi_{hi}$ 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.

6. A logistic regression model using two FFQs and 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 equation 4, by calibrating against the protein intake derived from till receipts, with proportions $\pi_{hi}$ 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 were available.

7. To explore the sensitivity of model 5 (point 5 above) to incomplete recording of intake by till receipts, on the basis of previous work (26, 27) we assumed that 12 percent of dietary intake was not captured by itemized till receipts but was recorded by means of 28-day shopping diaries that were subject to the same person-specific bias as the 24-hour recalls.

8. To explore the sensitivity of model 6 (point 6 above) to incomplete recording of intake by till receipts, we assumed that 12 percent of dietary intake was not captured by itemized till receipts but was recorded by means of 28-day shopping diaries that were subject to the same person-specific bias as the 24-hour recalls.

Ignoring measurement error more than halved the slope (log odds ratio) from 0.70 to 0.28, reducing an odds ratio from 2.0 to 1.3 (95 percent confidence interval: 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 being 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 led 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 that was consumed by each individual, $\pi_{hi}$, and the proportion of food purchased that was 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 that the proportions $\pi_{hi}$ and $c_h$ were known produced 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 use of diaries for calibration.
TABLE 1. Logistic regression coefficients $\hat{\alpha}_0$ and $\hat{\alpha}_1$, estimated reliability ratio for the food frequency questionnaire (FFQ), $\lambda_{FFQ}$, and correlation between the person-specific biases, $\rho(r,s)$, for eight different models of measurement error in dietary assessment*.

<table>
<thead>
<tr>
<th>Model</th>
<th>Components of model</th>
<th>Regression coefficient</th>
<th>Reliability ratio ($\lambda_{FFQ}$)</th>
<th>Correlation ($\rho(r,s)$)</th>
<th>Parameter estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Logistic regression ignoring measurement error</td>
<td>$-1.17 (0.76)$†</td>
<td>0.28 (0.19)</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>Repeat FFQ</td>
<td>$-1.39 (0.95)$</td>
<td>0.34 (0.24)</td>
<td>0.80 (0.02)</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>FFQ and recall</td>
<td>$-2.56 (1.63)$</td>
<td>0.60 (0.40)</td>
<td>0.32 (0.04)</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>Biomarker, FFQ, and recall</td>
<td>$-3.19 (1.22)$</td>
<td>0.74 (0.29)</td>
<td>0.37 (0.03)</td>
<td>0.23 (0.07)</td>
</tr>
<tr>
<td>5</td>
<td>Grocery till receipts, FFQ, and recall with $\pi_y$ and $\delta_y$ known (assuming that till receipts capture all dietary intake)</td>
<td>$-3.10 (1.18)$</td>
<td>0.69 (0.26)</td>
<td>0.29 (0.02)</td>
<td>0.42 (0.07)</td>
</tr>
<tr>
<td>6</td>
<td>Biomarker, till receipts, FFQ, and recall (assuming that till receipts capture all dietary intake)</td>
<td>$-3.16 (1.10)$</td>
<td>0.74 (0.26)</td>
<td>0.36 (0.03)</td>
<td>0.31 (0.08)</td>
</tr>
<tr>
<td>7</td>
<td>Till receipts, FFQ, and recall with $\pi_y$ and $\delta_y$ known (assuming that diaries are used to supplement till receipts for food consumed without receipts)</td>
<td>$-2.89 (1.08)$</td>
<td>0.71 (0.27)</td>
<td>0.29 (0.02)</td>
<td>0.40 (0.06)</td>
</tr>
<tr>
<td>8</td>
<td>Biomarker, till receipts, FFQ, and recall (assuming that diaries are used to supplement till receipts for food consumed without receipts)</td>
<td>$-3.02 (1.19)$</td>
<td>0.71 (0.28)</td>
<td>0.36 (0.03)</td>
<td>0.30 (0.08)</td>
</tr>
</tbody>
</table>

* True values used in simulations: $\hat{\alpha}_0 = -3$, $\hat{\alpha}_1 = 0.7$, $\lambda_{FFQ} = 0.3$, $\rho(r,s) = 0.3$, $\hat{\alpha}_n = 0.1$, $\hat{\beta}_{male} = 1$, $\hat{\beta}_{female} = 0.8$, and $\hat{\beta}_{child} = 0.5$.
† Numbers in parentheses, empirical standard deviation of the estimate.

DISCUSSION

The feasibility of collecting itemized grocery 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 manner similar to that of 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 being 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 percent of United Kingdom grocery shopping in 2000 was done in supermarkets (42), and many of the remaining smaller shops use itemized till receipts as well, in the same year 9 percent of the weekly food spending was done in restaurants and cafés (43), with only some establishments providing itemized bills. Consideration must therefore be given to meals eaten outside of the home, food purchased from shops that do not provide itemized receipts (such as staff canteens), and 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 information would be analyzed in the same way as a food diary to derive estimated amounts of nutrients based on standard portion sizes, which lack the precision of a weighed intake.

Use of any additional self-reported record of intake such as this reduces the objectivity of the method and introduces an unwanted element of person-specific bias, albeit less than that incurred with a wholly self-reported measure. Because of this, it is probably not appropriate to consider use of itemized till receipts a totally objective measure but rather to consider it more objective than use of food diaries or 24-hour recalls alone. However, in our study, 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 use of 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 these methods to be applied to a situation where no adequate biomarkers were available—for example, assuming that the same proportions $\pi_y$ hold across a range of different exposures. This implies that different
members of a household eat meals of identical nutritional content and that 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 will be required to tell whether 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 made for storage, such as the purchase of multipacks, food for home freezing, large containers of cooking oil, alcoholic beverages, etc., are characteristic of modern shopping habits; over half of United Kingdom consumers buy in bulk (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 taken 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 so on. 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 \( \sigma_i^2 \) were derived for receipts covering a particular time period, 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, data from biomarkers and reference instruments are only collected in a subsample of study participants, because of the cost. Use of biomarkers can also be invasive and requires a substantial amount of staff time for collecting and analyzing samples, while instruments such as weighed food diaries and 24-hour recalls require a substantial amount of 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 is used. It may be feasible to collect till receipts in a larger subsample than is possible with biomarkers, increasing the precision of the final estimate. Alternatively, till receipts might provide an appropriate instrumental variable with which 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 till receipts, which reduces the objectivity of the method. Using till receipts may require strong assumptions for deriving estimated intakes for individuals from household-level data, which weakens the method. Despite these substantial reservations, we propose that in situations where adequate biomarkers do not exist or are prohibitively expensive, 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, which are prone to large random variation in small validation samples.

ACKNOWLEDGMENTS

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

Conflict of interest: none declared.

REFERENCES


