UNDER- AND OVERREPORTING IN A FOOD FREQUENCY QUESTIONNAIRE

Julia L. Bienias, Rush-Presbyterian-St. Luke's Medical Center
Julia L. Bienias, Rush Inst. for Healthy Aging, 1645 W. Jackson Blvd., Suite 675, Chicago, IL 60612

KEY WORDS: Measurement error, Dietary data, Modeling, Aging, Questionnaire design, Cognitive aspects, Imputation

1. Overview

Food frequency questionnaires (FFQs) are commonly used to ascertain individuals' "usual" eating patterns. Respondents are asked to report the frequency with which they consume various food items and dietary supplements, based on the previous year's eating behavior. Intake of various micro- and macro-nutrients (e.g., Vitamin E, fat) are estimated based on a tabulation of the food frequencies. These nutrients are used to examine nutrient-disease (or other outcome) relationships. Responses to an FFQ typically reflect more intake than is actually consumed, but that bias may be differential: It has been suggested that misreporting may vary by type of person and type of food, for example, obese people may overreport fruit and vegetable intake and underreport fat intake.

We wished to investigate the applicability of these assertions to an elderly population and the impact that such hypothesized misreporting might have on the conclusions drawn from the data. In the present work, we used simulated data to examine the effect that such misreporting might have for two specific data-analytic models used to characterize the association between nutrient intake and physical activity and ability. Future directions include using the results of a calibration study that is currently in progress to look for possible evidence of misreporting, examining other possible misreporting mechanisms, generalizing the results based on the distributional properties of the predictor variables, and looking at methods of adjusting for the misreporting at the level of the question, based on an understanding of the cognitive mechanisms underlying the errors (versus adjusting the regression estimates *post hoc*, for example).

2. The Chicago Health and Aging Project

2.1 The General Study

The Chicago Health and Aging Project (CHAP) is a community-based prospective longitudinal cohort study of persons 65 and older in three neighborhoods on the south side of Chicago. The participants are racially and economically diverse, providing a reasonably representative urban population. The focus of CHAP is common health problems of the elderly, particularly prevalence and incidence of Alzheimer's disease and risk factors associated with it. For the initial data collection period (first wave), the CHAP staff conducted a door-to-door census to develop a frame of persons 65 and older; this was followed by a baseline interview of all such persons (N=6162) and then a stratified random sample of baseline participants for further clinical evaluation, including a structured uniform clinical evaluation to determine whether or not the participant was likely to have Alzheimer's disease. Coverage and response rates to the survey were quite high, with less than 1% estimated undercoverage in the census and over 80% response rate for the baseline interview. The plan is for the full population interview to be repeated every 3 years, with each wave having additional sample drawn for detailed clinical evaluation, in addition to having people "age into" the study. (See Wilson et al., 1999, for a further description of the study.)

2.2 The Nutrition Component

In addition to the main CHAP study, a nutritional component was added to investigate dietary risk factors in chronic health problems common among older persons. Dietary information was collected via a modified Willett Semi-Quantitative Food Frequency Questionnaire (FFQ; Willett et al., 1985; Willett, 1998). The mode of administration was mail-out/mail-back, with telephone and interviewer follow-up. When necessary, field personnel administered the FFQ as an interview by

---

We graciously thank Denis Evans and Martha Clare Morris for the use of their data, which formed the basis for the simulation; Walter Willett for generously sharing his program to compute nutrients from the food frequency questionnaire; Helaine Rockett and Steven Stuart for their assistance with using the program; Martha Clare Morris, Bob Wilson, Christy Tangney, and Laurel Beckett for helpful discussions; and Woojeong Bang for help with programming. The views expressed are attributable to the author and do not necessarily reflect those of the Rush Institute for Healthy Aging. The author can also be contacted via e-mail at *jbienias@rush.edu*.

reading the questions directly from the questionnaire. All participants in the main CHAP baseline full-population interview were invited to participate in this additional study. FFQs were obtained from 94% of surviving respondents. (A separate calibration study is currently ongoing, which will be comparing the FFQ data to an average of six 24-hour dietary recalls, as well as to levels of some nutrients estimated from blood samples.) FFQs will be administered in conjunction with future waves of the CHAP full-population interview.

In this paper, we used data from the initial wave of full-population data collection as the basis for our simulations. FFQs lacking sufficient data to analyze (e.g., more than half of the food questions missing) were excluded from analysis, resulting in a total of 5017 respondents as of this writing (a few more questionnaires are still in the field). Tables 1 and 2 show the breakdown of these participants with respect to age, race, and education, where these variables were known. (Nearly all nonblacks were white.)

Table 1. Race & Age of Baseline Participants in Nutrition Component of CHAP

	Age		
	65-69	70-79	80+
Nonblacks	*463*	*972*	*532*
Blacks	*1119*	*1470*	*461*

Table 2. Education of Baseline Participants in Nutrition Component of CHAP

	Formal Education
Grade School	*865*
High School	*2294*
College +	*1840*

3. The Food Frequency Questionnaire

The Willett questionnaire was modified for use with this population; in particular, the questions were typeset into a larger and easier-to-read format; portion sizes were made non-quantitative (e.g., "1", "one slice"); shorter response options were listed directly below each item (vs. the grid pattern of the original questionnaire); certain whole dishes (e.g., macaroni and cheese) were included; and one local ethnic food item was added based on pilot testing (Morris, Colditz, & Evans, 1998).

Questions are grouped by food type into seven different categories (Beverages, Dairy, Main Dishes, Bread/Cereal, Fruits/Vegetables, Snacks, Misc. (e.g., condiments)), for a total of 121 substantive questions as well as questions about vitamin supplement use, frequency of eating out, and other administrative questions. Respondents are asked to report their frequency of consuming each food item over the past year, such as:

Cheeseburger (1)
- o Never/less than 1 per month
- o 1 -3 per month
- o One per week
- o 2 - 4 per week
- o 5 or more per week

The number of response choices varies by food item.

Based on the responses and assumed or specified portion sizes, nutrient intake is computed using a program developed by staff at the Harvard School of Public Health, where the original questionnaire was developed. Over eighty macro- (e.g., fat, protein) and micronutrient (e.g., vitamin B12) values are derived, including measurement of various nutrients including or excluding reported vitamin and mineral supplement intake.

4. Impact of Possible Differential Misreporting

4.1 Hypotheses Examined

In general, misreporting food intake on the FFQ could be caused by one or more of the following: lack of understanding of the questions or the food items mentioned; lack of access to the information in memory, either because it was not stored in the first place (e.g., because eating was seen as a mundane behavior, see Schwarz & Hippler, 1987) or because it cannot be retrieved; an error in determining the frequency with which a given food was eaten with respect to the reference period (for example, due to telescoping; e.g., see Sudman & Bradburn, 1974); and various cognitive mechanisms that might lead a respondent to misreport in particular ways. These latter are described in more detail in the Discussion. (See Strube, 1987, for an overview of the memory processes in answering survey questions, and Groves, 1989, for further discussion.)

In the present investigation, we focused on a particular type of misreporting. Previous research has suggested that people who are obese may misreport their intake; in particular, they may underreport their intake of fatty foods (e.g., Lichtman et al., 1992; Heitman & Lessner,

1995) and possibly overreport their fruit and vegetable intake. If true, one can postulate several mechanisms that might lead to this. For example, overweight people have been shown to be more tuned in to external cues when eating (Rodin, 1976; Rodin, 1980; Rodin, Slochower, & Fleming, 1977); it is possible this might inhibit them from forming clear memories about the frequency of their intake of different foods and thus lead them to misreport in ways they believe to represent "correct" eating. Note that to truly test for misreporting, one would need a "gold standard" measure of some sort; simply showing that obese people reported low frequencies of eating fatty foods compared to non-obese people, for example, would not "prove" anything: They may in truth have been eating less fat over the reference period. In the absence of that, we focused on a sensitivity-type approach, examining the impact of potential misreporting on analysis. To summarize the two specific hypotheses:

(a) Obese people will underreport their intake of fat by underreporting their intake of foods high in fat); and
(b) Obese people will overreport their intake of fruits and vegetables.

(Assumed here is that non-obese people will not show any differential misreporting.)

4.2 Method and Definitions

To examine the impact of the hypothesized misreporting, it was necessary to create data with particular reporting patterns. For the purposes of this initial research, we restricted ourselves to women from the baseline group discussed previously. Because the focus of the paper was on cognitive-based reporting errors, we excluded people with evidence of possible cognitive impairment, as evidenced by having a score on the Mini Mental State Examination (Folstein, Folstein, & McHugh, 1975) at or below 23. This yielded 2419 women who provided the basis for the simulations.

There are many definitions of "obese;" we used a common one based on the body mass index (BMI; Quetelet, 1869): obese if $Wt_{kg} / Ht_m^2 \geq 30$. We then simulated data that would represent an "adjustment back" for the different theorized error patterns and examined the impact of the different imputations on the model-fitting results.

Specifically, there were 32 food items listed in the section labeled "Fruits &Vegetables." Although some vegetables or fruits might be consumed in other foods (such as tomato sauce in lasagna), using only the 32 primary questions seemed a very reasonable approach to "adjust back for" possible overreporting of fruits and vegetables. For fatty foods, we considered the nine foods with the highest fat content: butter, margarine, cream cheese, whipped cream, potato chips, corn chips, peanuts, hot dogs, and salami, bologna or similar meat sandwiches; these foods had substantially higher fat content than other foods on the questionnaire.

To construct the simulated data sets, within the "obese" group, we either set all the responses for either category to the extreme value that would represent "adjusting back" the possible reporting error (e.g., by setting all the fruit and vegetable responses to "Never/less than once a year" or similar) or by choosing half of the people randomly and then randomly adjusting half of their values by 2 levels if possible. The exact adjustment depended on the number of choices for the question. For example, if there were 5 choices (1-5) for a fruit question and the response were a "4," an adjustment "down" would be setting that response to a "2." If the response were a "2," the adjustment would be to a "1," which is the lowest possible value. Ten iterations were performed for each of the random-sample-based simulations and the results averaged; the variance among the estimates was extremely low, justifying the use of such a small number of iterations.

4.3 Results

We first examined whether or not there appeared to be evidence of differential reporting for women with low BMI versus those with high BMI. Because one of the two parts of the hypothesis was that being obese would lead to differential underreporting of fat intake, we first examined the reported percent calories from fat. The first two columns of Table 3 show the number of women broken down by BMI grouping and reported percent calories from fat ($\chi^2_2 = 2.371$, $p = 0.306$). Without a "gold standard," there is no way to ascertain whether or not there is differential misreporting.

Table 3. Percent Calories from Fat

% cal. from fat	BMI 0-30	BMI 30+		
	Actual	Actual	fruit/veg ↓	fat ↑
≤ 20%	*59*	*26*	*3*	*0*
20-30%	*668*	*289*	*113*	*0*
30% +	*838*	*314*	*513*	*629*

Next, to demonstrate the impact of the hypothesized reporting errors, we computed the percent calories from fat for two of the extreme simulated data sets. The third column in Table 3 shows the tabulation of women with high BMI whose data were "adjusted back" to report the absolute minimum intake of fruits and vegetables. The fourth column shows the simulation wherein the nine fatty foods were "adjusted back" to the maxiumum possible frequency. As expected, this last column shows everyone with high BMI in the high-fat-intake cell; similar shifting can be seen for the fruit/vegetable simulation, because the dietary composition has changed.

To assess the impact of possible misreporting on the conclusions drawn from an analytic model, we considered two reasonable models of interest:

- Logit(any activity) = f(saturated fat, physical function, age, race, education), where "any activity" was a binary variable coded "1" if the person reported engaging in any of a number of recreational physical activities such as walking or dancing; physical function was based on the Nagi scale (Nagi, 1976), a measure of five basic motor skills (e.g., bending, pushing or pulling a large object) -- the variable was the number of items a person reported being able to do; race was a binary variable (black/non-black); education was years of education.

- Physical Function = f(carbohydrates, age, race), where physical function was again measured by the Nagi. This was fit using ordinary least squares.

Both saturated fat and carbohydrates were adjusted for overall energy intake ("calorie-adjusted") based on the regression approach described in Willett (1998), which yields a purer estimate of intake that reflects dietary composition (versus absolute intake). It is important to do this adjustment with FFQs, because respondents generally overreport intake with this type of measure. Physical function and education were included as additional covariates in the logistic regression model, because: (1) Engaging in physical activity would depend on one's level of functioning, and (2) Choosing to engage in physical activity if one is able may be affected by education. (Remember, sex is not included because the dataset includes only women.)

The results are presented in Table 4. The rows are either the estimated odds ratio for saturated fat for the first model or the *p*-value associated with the carbohydrate term for the second model; both of these are easier to interpret than the raw coefficients. The data columns are defined as follows:

(1): The actual data as reported;

(2): The data based on imputing the obese women to have minimal fruit/vegetable intake;

(3): The data based on the average of the ten random-sample-based simulations, changing fruit/vegetable questions only;

(4): The data based on imputing the obese women to have maximal intake of the nine fatty foods (but not adjusting their fruit/veg. intake);

(5): The data based on the average of the ten random-sample-based simulations, changing the nine fatty food questions only; and

(6): The combination of (2) and (4): Setting the intake of all fruits/vegetables to minimum and all of the nine fatty foods intake to maximum.

Table 4. Results from Sample Analyses

	(1)	(2)	(3)	(4)	(5)	(6)
OR for sat. fat.	*.96*	*.95*	*.96*	*.97*	*.90*	*.98*
p for carb.	*.89*	*.03*	*.81*	*.02*	*.72*	*.02*

As can be seen, the extreme imputations, representing maximal misreporting (columns 2, 4, and 6) have a strong impact for the ordinary least squares regression model but minimal impact for the logistic regression model. The more realistic simulated data (columns 3 and 5) do not appear to change the conclusions much, as compared to the actual data. That is, if half the high BMI respondents were misreporting half their fruit and vegetable or fatty food intake, and the data were "adjusted back," the conclusions would not change (for this example).

5. Discussion

Three hypotheses can be advanced regarding the cognitive basis for the possible misreporting mechanism examined in this paper: impression management, demand characteristics of the experimental setting, and lack of attention to food in general leading to a lack of stored information about one's eating habits. In the case of the latter, one might not expect differential misreporting, but either simply random errors or perhaps mostly uniform under- or overreporting relative to someone of normal weight. Clearly these mechanisms could be connected: For example, if one has minimal access to true information but is focused on impression management, the responses would reflect one's beliefs about "good eating" and possibly have differential errors.

Various writers, beginning with Bingham and Moore (1959) have discussed the interview as a type of social interaction, a "conversation with a purpose" (e.g., see also Sudman & Bradburn, 1982). The implications of this are that both the norms of a social interaction and those of an experimental setting may apply to a given interviewing situation. Thus, a respondent may respond in ways that enhance his or her self-presentation (impression management) and/or in ways that meet the needs of the interviewer as perceived by the respondent (demand characteristics). (For further discussion of self-presentation issues, see Fiske & Taylor, 1991; Snyder, 1987; Baumeister, 1986).

Regarding both impression management and demand characteristics, one of the limitations of the foregoing (aside from being a limited empirical study) is that the differential misreporting discussed assumes the respondent has some understanding of the nutritional or fruit/vegetable content of the foods listed on the FFQ. In the case of fruits and vegetables, this assumption seems amply justified; if nothing else, the questions we considered were in a section labeled "Fruits & Vegetables." However, in the case of the possible misreporting of one's intake of fatty foods, it is certainly quite possible respondents might not know exactly how high in fat were the foods we considered. The reason for choosing such a short list (nine) of high-fat foods was to keep the fat content extremely high relative to the other foods in the FFQ; that way, we hoped that even though a respondent might not know the exact fat content, they might know these foods were considered "high." For impression management to operate, respondents must have some knowledge of the "value" of different foods (e.g., that hot dogs are high in fat, and they do not wish, consciously or unconsciously, to be seen as someone who eats foods high in fat). And, if demand characteristics are even partly operable here, respondents must have a sense of what the researcher wants from them, which might also entail knowing the nutritional value of the foods.

Thus, if there is differential reporting by level of BMI, say, the observed effect of it might be mediated by the level of knowledge of the respondent, as well as by his/her need for impression management or need to meet the demands of the research setting. For example, it would be interesting to examine some of these potentially competing cognitive mechanisms in conjunction with a measure of public self-consciousness (Fenigstein, Scheier, & Buss, 1975; Carver & Scheier, 1981) and a measure of nutrition knowledge (administered after the FFQ, needless to say!). Finally, because the questions on the FFQ require the respondent to choose a frequency of eating the food from among a given set of categories, one could further explore the interaction between the response alternatives and self-consciousness (Schwarz & Bienias, 1990).

The approach to measurement errors due to misreporting of food intake taken in this paper differs considerably from the more common model-based approaches (e.g., Willett, 1998, Ch. 12; Rosner, Spiegelman, & Willett, 1990; Freedman, Carroll, & Wax, 1991; see also Biemer et al., 1991). These approaches use the estimated relation between the primary nutrient estimate (e.g., from an FFQ) and a "better" measure (e.g., from a blood sample) to produce a factor that is used to adjust the odds ratios or other model estimates. Such models can be made more complex to incorporate systematic misreporting (e.g., Prentice, 1996). The current paper seeks not to adjust the final model but to examine the mechanism underlying errors. (See Groves, 1999, for another very recent treatment outlining similar goals.) Ideally, if we could predict with accuracy the types of errors people make, we could adjust the data <u>at the item level</u>, which would be the most general type of adjustment we could make.

6. References

Baumeister, R. F. (Ed.). (1986). *Public Self and Private Self.* NY: Springer-Verlag.

Biemer, P. P., Groves, R. M., Lyberg, L. E., Mathiowetz, N. A., & Sudman, S. (Eds.). (1991). *Measurement Errors in Surveys*. NY: Wiley.

Bingham, W. V. D., & Moore, B. V. (1959). *How to Interview* (4th ed.). NY: Harper & Row.

Carver, C. S., & Scheier, M. F. (1981). *Attention and Self-Regulation: A Control-Theory Approach to Human Behavior*. NY: Springer-Verlag.

Fenigstein, A., Scheier, M. F., & Buss, A. H. (1975). Public and private self-consciousness: Assessment and theory. *Journal of Consulting and Clinical Psychology*, **43**, 522-527.

Fiske, S. T., & Taylor, S. E. (1991). *Social Cognition* (2nd ed.). NY: McGraw-Hill.

Folstein, M. J., Folstein, S. E., & McHugh, P. R. (1975). Mini mental state: A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, **12**, 189-198.

Freedman, L. S., Carroll, R. J., & Wax, Y. (1991). Estimating the relation between dietary intake obtained from a food frequency questionnaire and true average intake. *American Journal of Epidemiology*, **134**, 310-320.

Groves, R. M. (1989). *Survey Errors and Survey Costs.* NY: Wiley.

Groves, R. M. (1999). Survey error models and cognitive theories. In Sirken, M. G., Herrmann, D. J., Schechter, S., Schwarz, N., Tanur, J. M., & Tourangeau, R. (Eds.), *Cognition and Survey Research.* NY: Wiley, 235-250.

Heitmann, B. L., & Lessner, L. (1995). Dietary underreporting by obese individuals -- is it specific or non-specific? *British Medical Journal*, **311**, 986-989.

Johansson, L., Solvoll, K., Bjørneboe, G.-E. A., & Drevon, C. A. (1998). Under- and overreporting of energy intake related to weight status and lifestyle in a nationwide sample. *American Journal of Clinical Nutrition*, **68**, 266-274.

Lichtman, S. W. , Pisarska, K., Berman, E. R., Pestone, M., Dowling, H., Offenbacher, E., Weisel, H., Heshka, S., Matthews, D. E., & Heymsfield, S. B. (1992). Discrepancy between self-reported and actual caloric intake and exercise in obese subjects. *New England Journal of Medicine*, **327**, 1893-1898.

Morris, M. C., Colditz, G. A., & Evans, D. A. (1998). Response to a mail nutritional survey in an older bi-racial community population. *Annals of Epidemiology*, **8**, 342-346.

Nagi, S. Z. (1976). An epidemiology of disability among adults in the United States. *Milbank Memorial Fund Quarterly*, **54**, 439-468.

Quetelet, L. A. (1869). *Physique Sociale.* Brussels: C. Muquardt.

Rodin, J. (1976). The role of perception of internal and external signals on the regulation of feeding in overweight and nonobese individuals. In T. Silverstone (Ed.), Appetite And Food Intake. Braunschweig: Pergamon Press/Vieweg, 265-283.

Rodin, J. (1980). Social and immediate environmental influences on food selection. International Journal of *Obesity*, **4**, 364-370.

Rodin, J., Slochower, J., Fleming, B. (1977). Effects of degree of obesity, age of onset, and weight loss on responsiveness to sensory and external stimuli. *Journal of Comparative and Physiological Psychology*, **91**, 586-597.

Rosner, B., Spiegelman, D., & Willett, W. C. (1990). Correction of logistic regression relative risk estimates and confidence intervals for measurement error: The case of multiple covariates measured with error. *American Journal of Epidemiology*, **132**, 734-745.

Schwarz, N., & Bienias, J. (1990). What mediates the impact of response alternatives on frequency reports of mundane behaviors? *Applied Cognitive Psychology.* **4**, 61-72.

Schwarz, N., & Hippler, H.-J. (1987). What response scales may tell your respondents. In Hippler, H.-J., Schwarz, N., & Sudman, S. (Eds.), *Social Information Processing and Survey Methodology.* NY: Springer-Verlag, 163-178.

Snyder, M. (1987). *Public Appearances, Private Realities.* NY: W. H. Freeman.

Strube, G. (1987). Answering survey questions: The role of memory. In Hippler, H.-J., Schwarz, N., & Sudman, S. (Eds.), *Social Information Processing and Survey Methodology.* NY: Springer-Verlag, 86-101.

Sudman, S., & Bradburn, N. M. (1974). *Response Effects in Surveys.* Chicago: Aldine.

Sudman, S., & Bradburn, N. M. (1982). *Asking Questions.* San Francisco, CA: Jossey-Bass.

Willett, W. C. (1998). *Nutritional Epidemiology* (2nd ed.). NY: Oxford Univ. Press.

Willett, W. C., Sampson, L., Stampfer, M. J., Rosner, B., Bain, C., Witschi, J., Hennekens, C., Speizer, F. E. (1985). Reproducibility and validity of a semi-quantitative food frequency questionnaire. *American Journal of Epidemiology*, **122**, 51-65.

Wilson, R. S., Bennett, D. A., Beckett, L. A., Morris, M. C., Gilley, D. W., Bienias, J. L., Scherr, P. A., & Evans, D. A. (1999). Cognitive activity in older persons from a geographically defined population. *Journal of Gerontology: Psychological Sciences*, **54B**, P155-P160.