Article

Impact of identifying plausible respondents on the under-reporting of energy intake in the Canadian Community Health Survey

by Didier Garriguet

October, 2008

Canada

Impact of identifying plausible respondents on the under-reporting of energy intake in the Canadian Community Health Survey

by Didier Garriguet

Abstract

Background
Under-reporting is common in nutrition surveys. The identification of plausible respondents is a way of measuring the impact of under-reporting on the relationship between energy intake and body mass index (BMI).

Data and methods
A 24-hour dietary recall from 16,190 respondents aged 12 or older to the Canadian Community Health Survey (CCHS) - Nutrition was used to determine energy and nutrient intake. To identify plausible respondents, a confidence interval was applied to total energy expenditure derived from equations developed by the Institute of Medicine. Estimates of energy and nutrient intake for plausible respondents were compared with estimates for all respondents. Linear regression was used to demonstrate the impact of under-reporting on the relationship between reported energy intake and weight. Logistic regression was used to determine the impact of under-reporting on modelling the characteristics of obese people.

Results
With a confidence interval of 70% to 142% around energy expenditure, 57% of CCHS respondents were identified as "plausible respondents." Nutrient under-reporting varied between 1% and 10%. Analysis based only on plausible respondents re-establishes the theoretical relationship between energy intake and body weight, a relationship that is lost when analysis is based on the full sample.

Interpretation
Identifying plausible respondents is an effective way of measuring the impact of under-reporting food intake. Conclusions based on plausible respondents, rather than on all respondents, are more in line with theoretical expectations, such as a positive association between high energy intake and obesity.

Keywords
Caloric intake, diet, food habits, energy expenditure, energy metabolism, nutrition surveys, obesity, twenty-four-hour recall, under-reporting

Author
Didier Garriguet (1-613-951-7187; didier.garriguet@statcan.ca) is with the Health Information and Research Division at Statistics Canada, Ottawa, Ontario, K1A 0T6.

The 2004 Canadian Community Health Survey (CCHS)–Nutrition was the first national survey of the eating habits of the Canadian population since the early 1970s. One of the objectives of the 2004 CCHS was to determine the intake of energy (calories), macronutrients (fats, proteins and carbohydrates) and micronutrients (vitamins and minerals) for different groups.

While every effort was made to ensure that the data were accurate—from questionnaire design through sampling and interviewer training to raw data verification and validation—the CCHS, like most nutrition surveys, was subject to under-reporting.[1-6] For a number of reasons (forgetfulness, social desirability, self-image, fear of being negatively judged[5,6]), respondents may deliberately or inadvertently report that they ate and drank less than they actually did.

The CCHS used a well-established collection instrument, the Automated Multiple-Pass Method (AMPM),[7,8] to maximize respondents' recall of what they consumed the day before they were interviewed. However, as reported in a companion article,[9] under-reporting among the population aged 12 or older still amounted to an average of about 10% of total energy intake. Under-reporting was associated with a number of characteristics, notably, body mass index, age, sex, and physical activity.

The under-reporting of energy intake and of specific nutrients has implications for the analysis of CCHS data. For instance, relationships between the amount and types of food consumed and body mass index are obscured by under-reporting. This article aims to address such issues by identifying "plausible" respondents.

Methods

Data source

The 2004 Canadian Community Health Survey (CCHS)–Nutrition collected information about the food and nutrient intake of the household population at the national and provincial levels. It excluded members of the regular Canadian Forces; residents of the three

territories; people living on Indian reserves, in institutions, and in some remote areas; and all residents (military and civilian) of Canadian Forces bases. Detailed descriptions of the CCHS design, sample and interview procedures are available in a published report.[10]

A total of 35,107 people completed an initial 24-hour dietary recall; a subsample of 10,786 completed a second recall three to ten days later. Response rates were 76.5% and 72.8%, respectively. The energy and nutrient composition of the foods reported during each recall was determined according to Health Canada's Canadian Nutrient File (2001b Supplement).[11]

The original intention of the CCHS was to weigh and measure all respondents aged 2 or older, but for various reasons, the weight and height of around 40% of them were not measured. To adjust for this non-response, another survey weight was created, based on respondent classes with similar socio-demographic characteristics. Because of the bias that has been observed between self-reported and measured data,[12,13] measured height and weight are preferable to self-reports. Therefore, respondents with measured height and weight data (with the appropriate survey weight) were used for this analysis.

This study pertains to 16,190 respondents aged 12 or older who answered the physical activity questions. Women who were pregnant or breastfeeding, people of very low weight (body mass index less than 18.5kg/m^2), and respondents with no or invalid dietary intakes were excluded.

Identifying plausible respondents

Identifying plausible respondents requires establishing lower and upper cut-offs for their total predicted energy expenditure; that is, a range for the amount of energy they could be expected to expend to remain at their measured weight. The total predicted energy expenditure of CCHS respondents was determined with the equations developed by the Institute of Medicine (IOM),[14] which model the energy expenditure derived from a doubly labelled water study based on age, height and weight (body mass index), and level of physical activity. Details about the derivation of physical activity from the CCHS and about the IOM equations are presented in the accompanying article on energy under-reporting.[9]

Every CCHS respondent was identified as a plausible respondent, an under-reporter or an over-reporter, based on a comparison of their total predicted energy expenditure with their reported energy intake. Goldberg et al.[15] were the first to suggest such an approach, by creating a confidence interval for physical activity level (PAL) based on coefficients of variation (CV) of subjects' energy intake (CV_{wEi}), the accuracy of the measurement of their basal metabolic rate (CV_{wB}), and the total variation in physical activity level (CV_{tP}). Black[16] developed a practical guide for using the cut-offs, and explained the method's limitations. McCrory et al. went further with a direct comparison of total predicted energy expenditure and measured energy intake. In an initial study,[17] the model of total energy expenditure used a limited database of only 93 individuals, whereas a second[18] used the IOM equations, which were developed with information from more than 700 individuals. Both cases assumed the "low active" physical activity category for all individuals. In the report, *What America Drinks*,[19] McCrory's method was modified to produce larger intervals for plausible intakes by assuming four different levels of physical activity for every individual.

For the present analysis, McCrory's intervals for the four levels of physical activity were applied to CCHS respondents according to the amount of activity each of them reported. That is, the interval applied to CCHS respondents depended on whether they were sedentary, low active, active or very active.

Table 1
Estimation of standard deviation (SD), by age group and sex, household population aged 12 or older, 2004

Age group and sex	Size of double-recall sample	CV_{rEI}^2	CV_{pER}^2	CV_{mTEE}^2	SD
Total	5,581	32.6	10.6	8.2	35.3
12 to 13					
Male	298	30.3	5.4	8.2	31.8
Female	257	28.3	7.4	8.2	30.4
14 to 18					
Male	566	33.2	4.8	8.2	34.5
Female	583	33.6	7.2	8.2	35.4
19 to 30					
Male	402	36.0	10.4	8.2	38.3
Female	552	35.9	10.5	8.2	38.3
31 to 50					
Male	410	31.9	11.0	8.2	34.7
Female	512	34.2	10.8	8.2	36.8
51 to 70					
Male	434	32.8	11.7	8.2	35.8
Female	619	33.3	11.6	8.2	36.2
71 or older					
Male	331	28.4	13.3	8.2	32.4
Female	617	29.1	13.4	8.2	33.0

Note: CV_{rEI}^2 is within-individual variation in reported energy intake. CV_{pER}^2 is within-individual variation in predicted energy requirement. CV_{mTEE}^2 is within-individual variation in measurement error and day-to-day variation in total energy requirement (TEE).
Sources: 2004 Canadian Community Health Survey - Nutrition.

Statistics Canada, Catalogue no. 82-003-XPE • Health Reports, Vol. 19, no. 4, December 2008
Impact of identifying plausible respondents • Methodological Insights

3

The confidence interval for the ratio of measured energy intake (rEI) to the predicted energy requirement (pER) was constructed from the standard deviation (SD), defined as follows:

$$SD = \sqrt{CV_{rEI}^2 / d + CV_{pER}^2 + CV_{mTEE}^2}$$

where CV_{rEI}^2 represents the intra-individual variation of energy intake; d the number of days of recall; CV_{pER}^2, the error in predicted energy requirements; and CV_{mTEE}^2, the day-to-day variation and the measurement error for total energy expenditure based on doubly labelled water.

Black and Cole[20] estimated CV_{mTEE}^2 at 8.2%, which was used in the present study. CV_{rEI}^2 and CV_{pER}^2 came from the CCHS data. CV_{rEI}^2 came from the respondents who provided two dietary recalls, based on the formula:

$$CV_{rEI}^2 = \sqrt{\sum_{i=1}^{n} (CV_i^2) / n}$$

where CV_i is the CV calculated for every individual. CV_{pER}^2 was obtained by dividing the average standard error of individual predictions for a group by the average prediction of energy expenditure for that group.

The CCHS obtained two dietary recalls for approximately 30% of the sample, but only the first is used in the subsequent analysis. Therefore, $d=1$, and an average SD value of 35% is used (Table 1).

Because the energy intake distribution is skewed, the confidence intervals were constructed in the log scale, and the cut-offs were exponentiated. The confidence interval for the energy intake to energy expenditure ratio (EI:EE) for plausible respondents is

$$EI:EE \in [exp(-\alpha*SD); exp(\alpha*SD)]$$

A multiplicative factor α can be applied to the SD to construct the confidence interval. This study uses only the multiplicative factor of 1.

Respondents whose reported energy intake was less than 70% of their predicted energy expenditure were classified as under-reporters; if the figure was more than 142% of their predicted energy expenditure, they were classified as over-reporters. Plausible respondents were those whose energy intake was 70% to 142% of their predicted energy expenditure. The representativeness of this sample of plausible respondents was assessed by comparing their socio-demographic characteristics with those of the total sample.

Analytical techniques

When plausible respondents had been identified, under-reporting of energy and nutrient intake was determined by dividing estimates for plausible respondents by estimates for all respondents.

Linear regression was used to demonstrate the impact of under-reporting on the relationship between reported energy intake and weight for the total population and for plausible respondents. Logistic regression was used to determine the impact of under-reporting on modelling the characteristics of obese people. The bootstrap method, which takes the complex design of the CCHS into consideration,[21-23] was used to estimate the confidence intervals of estimated ratios and odd ratios. The significance level was set at p < 0.05.

Definitions

Three types of covariates were included in this study: lifestyle risk factors, health status, and socio-demographic characteristics.

Body mass index (BMI) is calculated by dividing weight in kilograms by the square of height in metres. In this analysis, the BMI categories for adults were based on Health Canada guidelines.[24] People whose BMI was between 18.5 kg/m² and 24.99 kg/m² were normal weight; between 25

Table 2
Slope of weight variable in modelling predicted energy expenditure or energy intake for all respondents and for plausible respondents, by age group and sex, household population aged 12 or older, Canada excluding territories, 2004

Age group and sex	Predicted energy expenditure model	Energy intake model	
		All respondents	Plausible respondents
12 to 13			
Male	27.01	2.34 *	19.97*
Female	18.52	-13.13 *	6.13*
14 to 18			
Male	24.94	-1.43 *	20.43*
Female	17.84	-6.63 *	7.97*
19 to 30			
Male	16.24	-9.25 *	12.58
Female	13.88	2.58 *	11.14
31 to 50			
Male	17.20	4.01 *	12.22*
Female	13.15	1.23 *	9.58
51 to 70			
Male	16.16	-1.87 *	9.57*
Female	13.15	-2.30 *	7.19*
71 or older			
Male	16.99	5.20 *	19.53
Female	13.43	-2.82 *	9.95*

* significantly different from slope estimate of predicted energy expenditure model (p < 0.05)
Note: Ratio of energy intake (EI) to energy expenditure (EE) of plausible respondents is between 0.70 and 1.42.
Source: 2004 Canadian Community Health Survey – Nutrition.

kg/33m² and 29.99 kg/m², overweight; and more than 30 kg/m², obese. For adolescents aged 12 to 17, the categories defined by Cole *et al.*[25] were used.

Four levels of *leisure-time physical activity* were determined: sedentary, low active, active, and very active.

Alcohol consumption refers to the 12 months before the CCHS interview.

Fruit and vegetable consumption was based on the reported usual frequency of consumption, not the 24-hour recall. It represents the number of times per day respondents consumed fruit and vegetables, not the amount of food consumed.

Smokers are those who reported that they smoke daily or occasionally.

The variables related to health status were *self-reported health* (excellent, very good, good, fair and poor) and the presence of at least one *chronic condition*.

The socio-demographic variables were: *sex* and *age*, based on the IOM dietary reference groups; *highest level of education in the household* (less than secondary graduation, secondary graduation, some postsecondary, postsecondary graduation); *household income* from all sources and accounting for household size (low, low/middle, middle, middle/high, and high); *employment status* the week before the interview; *immigrant* and *Aboriginal* status; and *province of residence*.

Results

One-third under-report

If an analysis uses only data for plausible respondents, the cost in terms of sample size may be high. Based on the confidence interval of 70% to 142% around the ratio of reported energy intake to predicted energy expenditure, 9,196 (57%) of CCHS respondents were identified as plausible respondents; 5,388 (33%) as under-reporters; and 1,606 (10%) as over-reporters.

The characteristics of plausible respondents did not differ significantly from those of the total population (Appendix Table A). However,

significant differences between plausible respondents and under- and over-reporters emerged in relation to BMI, physical activity, highest level of education in the household, and province. These differences persisted in a logistic regression model (data not shown).

Association between reported energy intake and weight

The biological relationship between energy intake and weight is obvious:

if weight is to be maintained, long-term energy expenditure must match long-term energy intake. The higher the weight, the higher the energy expenditure, and the greater the energy intake. Thus, in theory, the regression coefficients between weight and predicted energy expenditure requirements, and between weight and energy intake should be the same.

Table 2 shows the slope of weight in the model of total predicted energy *expenditure.* Because predicted energy

Table 3
Under-reporting of selected nutrients, household population aged 12 or older, Canada excluding territories, 2004

| Nutrients | Average or percentage | | Under-reporting | | |
	All respondents	Plausible respondents	% difference between plausible and all respondents	95% confidence interval from	to
Energy (kilocalories)	2,173	2,349	8.1	6.6	9.6
Proteins (grams)	87.6	92.6	5.7	3.9	7.6
% of calories	16.2	15.6	-3.3	-4.5	-2.2
Carbohydrates (grams)	270	291	8.0	6.4	9.5
% of calories	49.7	49.2	-0.9	-1.6	-0.3
Fats (grams)	79.4	86.8	9.3	7.4	11.3
% of calories	31.5	32.4	2.8	2.0	3.7
Monounsaturated fats (grams)	31.9	35.0	9.7	7.5	11.9
% of calories	12.6	13.0	3.7	2.7	4.7
Polyunsaturated fats (grams)	14.0	15.3	9.0	6.6	11.4
% of calories	5.6	5.7	2.8	1.4	4.1
Saturated fats (grams)	25.9	28.3	9.4	7.3	11.5
% of calories	10.3	10.6	2.7	1.4	3.9
Alcohol (grams)	9.2	10.1	8.8	1.1	16.6
Caffeine (milligrams)	210	211	0.6	-2.8	4.0
Water (grams)	2,765	2,833	2.4	1.2	3.7
Cholesterol (milligrams)	285	301	5.5	3.0	8.0
Dietary fibre (grams)	17.5	18.5	6.1	4.2	8.0
Sugar (grams)	111	121	9.6	7.8	11.5
Vitamin D (micrograms)	5.9	6.3	7.3	3.0	11.7
Vitamin C (milligrams)	131	140	6.9	4.3	9.4
Vitamin B12 (micrograms)	4.5	4.6	3.3	-1.6	8.2
Vitamin B6 (milligrams)	1.9	2.0	5.1	3.1	7.2
Vitamin A (micrograms)	713	759	6.4	2.5	10.4
Niacin (milligrams)	40.5	42.8	5.6	3.8	7.4
Riboflavin (milligrams)	2.0	2.1	7.2	5.6	8.8
Thiamin (milligrams)	1.8	1.9	6.8	4.9	8.8
Dietary folate equivalent (micrograms)	474	510	7.5	5.7	9.2
Folic acid (micrograms)	128	138	8.3	5.2	11.3
Naturally occurring folate (micrograms)	237	252	6.4	4.6	8.2
Calcium (milligrams)	907	983	8.3	6.4	10.2
Sodium (milligrams)	3,222	3,429	6.4	4.7	8.2
Iron (milligrams)	14.7	15.6	6.1	4.3	8.0
Magnesium (micrograms)	331	352	6.2	4.7	7.7
Phosphorus (milligrams)	1,387	1,479	6.7	5.0	8.3
Potassium (milligrams)	3,156	3,331	5.6	4.1	7.1
Zinc (milligrams)	11.6	12.3	5.7	3.7	7.7

Note: Ratio of energy intake (EI) to energy expenditure (EE) of plausible respondents is between 0.70 and 1.42.
Source: 2004 Canadian Community Health Survey – Nutrition.

Table 4
Ratio of energy intake and predicted energy expenditure of all respondents and plausible respondents, by body mass index (BMI) category, age group and sex, household population aged 12 or older, Canada excluding territories, 2004

	All respondents		Plausible respondents	
	Ratio EI:EE	Standard error	Ratio EI:EE	Standard error
Total	0.904	0.007	0.984	0.005
Body mass index category				
Normal	1.00	0.01	1.01	0.01
Overweight	0.87	0.01	0.97	0.01
Obese	0.79	0.01	0.96	0.01
Age group and sex				
12 to 13				
Male	1.01	0.03	1.02	0.01
Female	0.99	0.03	1.01	0.02
14 to 18				
Male	0.98	0.02	0.99	0.01
Female	0.95	0.02	1.00	0.01
19 to 30				
Male	0.96	0.02	0.98	0.01
Female	0.87	0.02	0.99	0.01
31 to 50				
Male	0.92	0.02	1.00	0.01
Female	0.88	0.02	0.97	0.01
51 to 70				
Male	0.88	0.02	0.99	0.01
Female	0.86	0.01	0.97	0.01
71 or older				
Male	0.84	0.02	0.96	0.02
Female	0.89	0.02	0.97	0.01

Note: Ratio of energy intake (EI) to energy expenditure (EE) of plausible respondents is between 0.70 and 1.42.
Source: 2004 Canadian Community Health Survey – Nutrition.

expenditure depends on body weight, the relationship would be expected to be strong. In fact, this is borne out, with the R^2 ranging from 0.51 to 0.77 (data not shown).

Table 2 also shows the slope of weight in the model of energy *intake* for all respondents and for plausible respondents. For all respondents, not only are all the slopes significantly different from the slope for total energy expenditure, but for 7 of the 12 age/sex groups, the slope is negative, indicating that the higher their measured weight, the lower their reported energy intake. In these models, R^2 never exceeds 0.04 (data not shown).

By contrast, for plausible respondents, the slope of weight is always positive. All slopes are closer to the theoretical biological relationship (the higher the weight, the greater the energy intake),

and there is no significant difference from the energy expenditure model for 4 of the 12 age/sex groups. Therefore, when based on plausible respondents, the quality of the energy intake model improves, with the R^2 ranging from 0.02 to 0.24 (data not shown).

Impact on reporting of nutrient intake

A comparison of the average reported consumption of a nutrient for all respondents with that for plausible respondents provides an estimate of the extent to which consumption of that nutrient is under-reported (Table 3). For example, based on the reported calorie intake of all respondents versus plausible respondents, the average rate of energy under-reporting is 8.1%, which is close to the 9.6% estimated

in the accompanying article.[9] Under-reporting of fat and sugar consumption amounts to 9.3% and 9.6%, respectively. Calcium (8.3%) and alcohol (8.8%) are also under-reported. (Negative values indicate that intake of the nutrient is over-reported by the total population.)

The ratio of reported energy intake to energy expenditure (EI:EE) is higher for plausible respondents (0.98) than for all respondents (0.90) (Table 4). Regardless of age and sex, the ratios for plausible respondents are higher, ranging from 0.96 to 1.02, compared with 0.84 to 1.01 for the total population. Even among obese people, who tend to under-report, the ratio is 0.96 for those identified as plausible respondents, compared with 0.79 for all respondents.

Impact on modelling obesity

The substantial under-reporting of food and beverage consumption in the CCHS has implications for analysis of the data. For example, the relationship between energy intake (calories consumed) and obesity among people aged 18 or older differs depending on whether the results are based on all respondents or on plausible respondents. Based on all respondents, no significant association emerges between calorie consumption and obesity (Table 5). However, the use of only plausible respondents yields a positive and significant association between obesity and calories consumed for both sexes.

Discussion

The main strength of this study is the large sample size. Even excluding under- and over-reporters, the number of plausible respondents is large enough and representative enough to permit a detailed analysis. As well, the availability of measured height and weight data obviated the need to account for the response bias associated with self-reports of these variables. Finally, the analysis incorporates a physical activity variable and considers several levels of activity when predicting energy expenditure requirements.

Table 5
Adjusted odds ratios relating energy intake and selected characteristics to obesity among all respondents and among plausible respondents, household population aged 18 or older, Canada excluding territories, 2004

	Male						Female					
	All respondents			Plausible respondents			All respondents			Plausible respondents		
		95% confidence interval			95% confidence interval			95% confidence interval			95% confidence interval	
Nutrients	Adjusted odds ratio	from	to	Adjusted odds ratio	from	to	Adjusted odds ratio	from	to	Adjusted odds ratio	from	to
Energy intake (in 100s of kcal)	0.99	0.98	1.01	1.10*	1.06	1.14	1.00	0.99	1.02	1.16*	1.12	1.20
Leisure-time physical activity												
Sedentary	1.20	0.80	1.79	1.21	0.67	2.19	1.61*	1.17	2.23	2.52*	1.55	4.10
Low active[†]	1.00	...	...	1.00	...	...	1.00	...	...	1.00	...	...
Active	0.86	0.47	1.58	0.59	0.22	1.63	0.39*	0.22	0.72	0.18*	0.06	0.51
At least one chronic condition												
Yes	1.61*	1.19	2.17	1.50	0.97	2.32	1.50*	1.18	1.92	1.38	0.92	2.08
No[†]	1.00	...	...	1.00	...	...	1.00	...	...	1.00	...	...
Education												
Less than secondary graduation	1.48	0.95	2.30	1.97*	1.09	3.55	1.36	0.99	1.89	1.59	0.96	2.66
Secondary graduation	1.05	0.72	1.53	1.42	0.85	2.38	1.36	0.90	2.07	1.22	0.68	2.18
Some postsecondary	1.14	0.69	1.89	1.48	0.71	3.09	2.32*	1.57	3.42	2.79*	1.48	5.26
Postsecondary graduation[†]	1.00	...	...	1.00	...	...	1.00	...	...	1.00	...	...
Self-reported health												
Excellent[†]	1.00	...	...	1.00	...	...	1.00	...	...	1.00	...	...
Good	1.30	0.87	1.94	0.89	0.51	1.56	1.45*	1.02	2.08	2.38*	1.42	4.00
Average	1.99*	1.28	3.08	1.74	0.92	3.30	2.50*	1.73	3.62	3.50*	1.98	6.19
Fair	2.76*	1.61	4.74	2.73*	1.19	6.26	2.77*	1.88	4.07	3.48*	1.85	6.56
Poor	3.39*	1.17	9.78	5.11*	1.38	19.01	2.46*	1.22	4.93	3.23*	1.17	8.89
Immigrant												
Yes	0.55*	0.38	0.81	0.60*	0.36	0.99	0.73	0.51	1.04	0.95	0.55	1.63
No[†]	1.00	...	...	1.00	...	...	1.00	...	...	1.00	...	...
Aboriginal												
Yes	1.25	0.57	2.70	1.29	0.44	3.77	1.92*	1.16	3.18	2.08	0.97	4.47
No[†]	1.00	...	...	1.00	...	...	1.00	...	...	1.00	...	...
Household income												
Low[†]	0.52	0.22	1.25	0.48	0.13	1.73	0.84	0.46	1.54	0.37*	0.17	0.79
Low/Middle	0.43*	0.20	0.90	0.39	0.14	1.07	0.88	0.53	1.47	0.86	0.39	1.93
Middle	0.76	0.48	1.20	0.95	0.51	1.78	1.21	0.84	1.75	1.30	0.74	2.28
Middle/High	0.90	0.65	1.25	1.03	0.64	1.67	1.04	0.75	1.45	1.05	0.63	1.75
High[†]	1.00	...	...	1.00	...	...	1.00	...	...	1.00	...	...
Age group												
18 to 30[†]	1.00	...	...	1.00	...	...	1.00	...	...	1.00	...	...
31 to 50	1.74*	1.23	2.45	2.56*	1.52	4.31	1.72*	1.23	2.39	2.14*	1.21	3.78
51 to 70	1.95*	1.34	2.84	3.22*	1.91	5.42	2.15*	1.52	3.05	3.11*	1.64	5.88
71 or older	0.85	0.52	1.39	1.66	0.78	3.52	1.19	0.80	1.77	2.17*	1.10	4.26

[†] reference category
* significantly different from estimate for reference category (p < 0.05)
... not aplicable
Note: Also adjusted for alcohol consumption in past 12 months, daily fruit and vegetable consumption, province, and smoking status. Categories for missing values for education and household income and the very active category for physical activity were included, but the odds ratios are not shown. The ratio of energy intake (EI) to energy expenditure (EE) for plausible respondents is between 0.70 and 1.42.
Source: 2004 Canadian Community Health Survey – Nutrition.

It would have been possible to assume, as was done by McCrory *et al.*[17] and Huang *et al.*,[18] that all respondents were "low active." This would result in a slightly smaller sample of plausible respondents (56%), but 96% of CCHS respondents would still be classified in the same category (under-reporter, over-reporter or plausible respondent) as in the present study. The difference lies in the representativeness of the sample, which would be slightly less if a single physical activity level had been used.

Another option for estimating physical activity would have been to use the method employed in *What America Drinks,*[19] whereby the lower confidence interval limit is set based on the assumption of a sedentary level of physical activity, and the upper

confidence interval limit is set assuming a very active level of physical activity. The result is that a much larger proportion of the CCHS sample— 74%—would be considered to be plausible respondents, including every plausible respondent identified in the present study. The drawback is that this method is less effective in correcting the distortion in the biological relationship between energy intake and body weight, and the relationship between energy intake and obesity in the logistic model is weaker.

An alternative is the Goldberg cut-offs for physical activity level (PAL) using the ratio of energy intake and predicted basal metabolic rates from the Schofield equations.[26] Using the same SD, with a basic comparative PAL of 1.55 and a multiplicative factor of 1, yields a slightly smaller sample of plausible respondents (54%) than in the present study. While the classification of respondents as under- and over-reporters and plausible respondents is the same for 90% of the sample, the representativeness of the plausible respondents is not as good.

The confidence intervals in this study were exponentiated, whereas McCrory *et al.*[17] and Huang *et al.*[18] used a ±35 % confidence interval. Application of the latter to the CCHS data would shift the confidence interval toward more under-reporters and fewer over-reporters, although 92% of respondents would still be classified in the same category as in the present study. The difference is in the level of correction of the energy intake to energy expenditure ratio (EI:EE). Instead of increasing the average EI:EE ratio for plausible respondents to 0.98, as is the case when using the log scale, a symmetrical confidence interval yields an average EI:EE ratio of 0.94.

Huang *et al.*[18] used virtually the same method as that employed in the present study to identify plausible respondents in the 1994-1996 Continuing Survey of Food Intakes by Individuals (CSFII). They identified plausible respondents through an optimal confidence interval

of ± 30.8% corresponding to 1.4 times the standard deviation (SD). McCrory *et al.*[17] used somewhat different predictive equations, but their confidence interval of ± 1 SD corresponded to ± 30%. In both cases, their confidence interval is similar to the one obtained in this analysis. The difference stems from the fact that the CSFII has a smaller CV for the energy intakes, compared with the Canadian population.

Limitations

A major limitation of the present study is that predicted energy expenditure requirements are based on a model constructed by the Institute of Medicine using a sample of about 700 people of different ages and levels of physical activity; very active people and some specific ages are under-represented in that sample.

The level of physical activity measured in the CCHS is self-reported. As well, it refers only to leisure time; work- or transportation-related physical activity is not included.

Excluding under- and over-reporters from an analysis also excludes some plausible respondents who happened to eat much more or much less than usual on the dietary recall day. Even so, on the basis of their socio-demographic characteristics, those who were identified as plausible respondents are representative of the total population.

Conclusion

It is essential to consider under-reporting in analyses of data from nutrition surveys, especially when examining relationships between diet and variables that are highly correlated with under-reporting, such as body mass index. This study shows that just over half of respondents to the Canadian Community Health Survey reported food and beverage consumption that was "plausible," given their height, weight, and level of physical activity. Fully a third were "under-reporters," in that what they reported that they ate and drank could not sustain their

measured weight. Under-reporting was particularly common among people who were obese, and as a consequence, under-reporting tends to distort analyses of the relationship between energy intake and obesity.

This study confirms the findings of others, specifically McCrory *et al.*[17] and Huang *et al.*,[18] on the impact of under-reporting on the biological relationship between energy intake and weight, as well as on models related to body mass index. And because of the large CCHS sample, these results can be confirmed for specific age and sex groups.

This study also validates the companion article on under-reporting.[9] The characteristics of under-reporters, compared with those of plausible respondents, are essentially the same as with the other method. Average energy under-reporting is slightly less in this analysis, but still in line with the results in the other article.

The technique employed in the present study has numerous applications. It can be used to analyze models related to body mass index and to estimate the level of under-reporting of nutrients, certain foods, and food groups. The results show that some foods are particularly susceptible to under-reporting. Future studies could attempt to better identify foods that are under-reported and occasions when the food consumed may not be reported. ■

References

1. Johnson RK, Soultanakis RP, Matthews DE. Literacy and body fatness are associated with underreporting of energy intake in US low income women using the multiple-pass 24-hour recall, a doubly labelled water study. *Journal of the American Dietetic Association* 1998; 98(10): 1136-40.

2. Jonnalagadda S., Mitchell DC, Smiciklas-Wright H, et al. Accuracy of energy intake data estimated by a multi-pass 24-hour dietary recall technique. *Journal of the American Dietetic Association* 2000; 100(3): 303-8.

3. Johansson G., Wikman A, Ahren AM, et al. Under-reporting of energy intake in repeated 24-hour recalls related to gender, age, weight status, day of interview, educational level, reported food intake, smoking habits and area of living. *Public Health Nutrition* 2001; 4(4): 919-27.

4. Pikholz C, Swinburn B, Metcalf P. Under-reporting of energy intake in the 1997 National Nutrition Survey. *Journal of the New Zealand Medical Association* 2004; 117(1202): 1-11.

5. Tooze JA, Subar AF, Thompson FE, et al. Psychosocial predictors of energy under-reporting in a large doubly labeled water study. *American Journal of Clinical Nutrition* 2004; 79: 795-804

6. Livingstone BE, Black AE. Markers of the validity of reported energy intake. *Journal of Nutrition* 2003; 133: 895S-920S.

7. Moshfegh AJ, Borrud L, Perloff B, et al. Improved method for the 24-hour dietary recall for use in national surveys. *The FASEB Journal: Official Publication of the Federation of American Societies for Experiment Biology* 1999; 13: A603 (abstract).

8. Moshfegh AJ, Raper N, Ingwersen L, et al. An improved approach to 24-hour dietary recall methodology. *Annals of Nutrition and Metabolism* 2001; 45(Supplement):156 (abstract).

9. Garriguet D. Under-reporting of energy intake in the Canadian Community Health Survey. *Health Reports* (Statistics Canada, Catalogue 82-003) 2008; 19(3): xx-xx.

10. Béland Y, Dale V, Dufour J, Hamel M. The Canadian Community Health Survey: Building on the success from the past. *Proceedings of the American Statistical Association Joint Statistical Meeting, Section on Survey Research Methods, August 2005.* Minneapolis: American Statistical Association, 2005.

11. Health Canada. *Canadian Nutrient File, 2005 Version.* Available at http://www.hc-sc.gc.ca/fn-an/nutrition/fiche-nutridata/index_e.html.

12. Tjepkema M. Adult obesity. *Health Reports* (Statistics Canada, Catalogue 82-003) 2006; 17(3): 9-25.

13. Shields M. Overweight and obesity among children and youth. *Health Reports* (Statistics Canada, Catalogue 82-003) 2006; 17(3): 27-42.

14. Institute of Medicine. *Dietary Reference Intakes for Energy, Carbohydrate, Fiber, Fat, Fatty Acids, Cholesterol, Protein and Amino Acids.* Washington, DC: National Academy Press, 2005.

15. Goldberg GR, Black AE, Jebb SA, et al. Critical evaluation of energy intake data using fundamental principles of energy physiology. 1. Derivation of cut-off values to identify under-recording. *European Journal of Clinical Nutrition*, 1991; 45: 569-81.

16. Black AE. Critical evaluation of energy intake using the Goldberg cut-off for energy intake: basal metabolic rate. A practical guide to its calculation, use and limitations. *International Journal of Obesity* 2000; 24: 1119-30.

17. McCrory MA, Hajduk CL, Roberts SB. Procedures for screening out inaccurate reports of dietary energy intake. *Public Health Nutrition* 2002; 5: 873-82.

18. Huang TT, Roberts SB, Howarth NC, McCrory MA. Effect of screening out implausible energy intake reports on relationships between diet and BMI. *Obesity Research* 2005; 13(7): 1205-17.

19. ENVIRON International Corporation. *What America Drinks.* Available at: http://www.thinkaboutyourdrink.com/pdf/What_America_Drinks_report.pdf.

20. Black AE, Cole TJ. Within- and between-subject variation in energy expenditure measured by doubly labelled water: implications for assessing the validity of reported dietary energy intake. *European Journal of Clinical Nutrition* 2000; 54: 386-94.

21. Rao JNK, Wu CFJ, Yue K. Some recent work on resampling methods for complex surveys. *Survey Methodology* (Statistics Canada, Catalogue 12-001) 1992; 18(2): 209-17.

22. Rust KF, Rao JNK. Variance estimation for complex surveys using replication techniques, *Statistical Methods in Medical Research* 1996; 5(3): 281-310.

23. Yeo D, Mantel H, Liu TP. Bootstrap Variance Estimation for the National Population Health Survey. *Proceedings of the Annual Meeting of the American Statistical Association: Survey Research Methods Section.* American Statistical Association: Baltimore, August 1999.

24. Health Canada. *Canadian Guidelines for Body Weight Classification in Adults* (Catalogue H49-179/2003E) Ottawa: Health Canada, 2003.

25. Cole TJ, Bellizzi MC, Flegal KM, et al. Establishing a standard definition for child overweight and obesity worldwide: international survey. *British Medical Journal* 2000; 32 (7244): 1240-3.

26. FAO/WHO/UNU. *Energy and protein requirements. Report of a Joint FAO/WHO/UNU consultation.*(Technical Report Series, 724). Geneva: World Health Organization, 1985.

Statistics Canada, Catalogue no. 82-003-XPE • Health Reports, Vol. 19, no. 4, December 2008
Impact of identifying plausible respondents • Methodological Insights

9

Appendix

Table A
Selected socio-demographic characteristics of plausible respondents, under-reporters and over-reporters, household population aged 12 or older, Canada excluding territories, 2004

Characteristics	All respondents	Plausible respondents	Under-reporters	Over-reporters
Average age (years)	42.9	42.6	44.5*	39.6*
		%		
Sex (female)	49.5	49.6	50.6	44.8
Body mass index category				
Normal	42.3	44.8	31.9*	61.4*
Overweight	35.3	35.1	37.8	29.0*
Obese	22.4	20.1	30.3*	9.6*
Leisure-time physical activity				
Sedentary	18.4	18.5	16.1	24.9*
Low active	74.5	75.2	75.0	69.5*
Active	6.6	5.9	8.2*	5.4E
Very active	0.5	0.5E	0.7E	F
Education				
Less than secondary graduation	9.8	9.1	11.1*	10.0
Secondary graduation	10.5	9.8	12.8*	7.6
Some postsecondary	7.2	7.1	8.0	5.3
Postsecondary graduation	70.3	71.8	66.1*	75.2
Daily fruit and vegetable consumption				
Fewer than 5 servings	68.0	67.2	70.8*	63.4
5 to 10 servings	30.4	31.3	27.6*	34.4
More than 10 servicing	1.3	1.0E	1.4E	F
Household income				
Low	2.6	2.1	3.3*	2.7
Low/Middle	5.6	5.4	6.2	4.8
Middle	19.3	19.0	20.3	17.3
Middle/High	33.1	33.3	31.5	37.4
High	30.4	31.4	29.6	27.3
Province				
Newfoundland and Labrador	1.7	1.7	1.9	1.3E
Prince Edward Island	0.4	0.4	0.5*	0.3E
Nova Scotia	3.0	3.2	2.9	2.1*,E
New Brunswick	2.4	2.5	2.5	1.7*
Quebec	24.0	25.8	18.3*	32.2*
Ontario	39.0	37.5	43.6*	32.0*
Manitoba	3.5	3.3	4.0*	2.9
Saskatchewan	2.9	2.8	3.3	2.5E
Alberta	9.8	9.2	10.8*	9.4
British Columbia	13.3	13.5	12.2	15.6

* significantly different from estimate for plausible respondents (p < 0.05)
... not aplicable
E use with caution (coefficient of variation 16.6% to 33.3%)
F too unreliable to be published (coefficient of variation greater than 33.3%)
Note: Ratio of energy intake (EI) to enegry expenditure (EE) of plausible respondents is between 0.70 and 1.42.
Source: 2004 Canadian Community Health Survey – Nutrition.