Interviewer effects in food acquisition surveys†

Ai Rene Ong[1],*, Mengyao Hu[1], Brady T West[1] and John A Kirlin[2]
[1]Institute for Social Research, University of Michigan, 426 Thompson Street, Ann Arbor, MI 48104, USA:
[2]US Department of Agriculture, Economic Research Service, Washington, DC, USA

Abstract

Objective: To understand the effects of interviewers on the responses they collect for measures of food security, income and selected survey quality measures (i.e. discrepancy between reported Supplemental Nutrition Assistance Program (SNAP) status and administrative data, length of time between initial and final interview, and missing income data) in the US Department of Agriculture's National Household Food Acquisition and Purchase Survey (FoodAPS).
Design: Using data from FoodAPS, multilevel models with random interviewer effects were fitted to estimate the variance in each outcome measure arising from effects of the interviewers. Covariates describing each household's socio-economic status, demographics and experience in taking the survey, and interviewer-level experience were included as fixed effects. The variance components in the outcomes due to interviewers were estimated. Outlier interviewers were profiled.
Setting: Non-institutionalized households in the continental USA (April 2012–January 2013).
Subjects: Individuals (*n* 14 317) in 4826 households who responded to FoodAPS.
Results: There was a substantial amount of variability in the distributions of the outcomes examined (i.e. time between initial and final interview, reported values for food security, individual income, missing income) among the FoodAPS interviewers, even after accounting for the fixed effects of the household- and interviewer-level covariates and removing extreme outlier interviewers.
Conclusions: Interviewers may introduce error in food acquisition survey data when they are asked to interact with the respondents. Managers of future surveys with similarly complex data collection procedures could consider using multilevel models to adaptively identify and retrain interviewers who have extreme effects on data collection outcomes.

Food acquisition surveys have been used to track habitual food and drink consumption, examine patterns of food acquisition by different demographic subgroups and understand food security[1–3]. Estimates produced by these surveys provide valuable insight for the drafting and modification of policies and programmes designed to promote healthy eating and better lifestyles, and for assisting groups of people who are identified as being food insecure.

A recent survey by the US Department of Agriculture (USDA), the National Household Food Acquisition and Purchase Survey (FoodAPS), represents such an effort to generate these important estimates[4]. FoodAPS is the first nationally representative survey in the USA to collect detailed week-long information on all household food acquisitions along with other related indicators such as food security, income, employment and participation in food assistance programmes. FoodAPS is a comprehensive survey that involved both face-to-face interviewing and a 7 d diary survey component to be completed by all members of the sampled households.

The data collection protocol for FoodAPS required an extensive amount of involvement from human interviewers. Interviewers were responsible for training the household to record their food acquisitions throughout the survey period using the diary. They were also responsible for conducting the initial and final face-to-face interviews. It is therefore important to understand if survey outcomes in FoodAPS (e.g. measures of food security, missing income reports) tend to have distributions that vary across interviewers, above and beyond any variance introduced by the quality of the instruments or the features of the respondents. Despite nearly a century of research into the

---

Corresponding author: Email aireneo@umich.edu

potentially negative effects of human interviewers on survey data collection[5], we are unaware of any prior research that has evaluated these effects in the context of food acquisition surveys. If these effects exist, they may have an impact on the conclusions drawn from using these data. The objective of the present study was to examine these systematic effects of interviewers on several different data collection outcomes in the FoodAPS.

Interviewer effects and FoodAPS

Researchers have long studied the variable effects of interviewers on different types of survey errors under the term 'interviewer effects'[5–12]. We define 'interviewer effect' here as the systematic effect of a particular interviewer on a given survey process (e.g. acquiring consent for record linkage) or outcome (e.g. insurance coverage, health). These effects may vary among different interviewers, giving rise to the term 'interviewer variance'. Interviewers have been found to use different techniques to list household members to select a respondent, as well as different techniques to identify and list occupied homes in a given area for sample selection[5,6]. This can introduce variable population coverage errors among interviewers if this listing is used to construct a sampling frame to draw samples from, as different interviewers will use different metrics to judge if a house is 'occupied'[6]. Interviewers have frequently been found to vary in terms of their achieved response rates[5] and some studies have presented evidence of interviewer variability in rates of consent for additional data collection or linkage of respondents' answers to other administrative data sources[7,8]. Finally, interviewers have frequently been found to introduce varying measurement errors, with subjective, sensitive and complex questions more prone to this effect than others[5,9].

These effects have generally been attributed to varying interviewer skills in convincing respondents to cooperate with a survey request; experience, mobility and attitudes of the interviewers towards the survey; and, somewhat inconclusively, sociodemographic characteristics[5,10]. Given that the FoodAPS relies heavily on interviewer involvement throughout the field period, it presents a unique opportunity to study the presence and magnitude of interviewer effects on key outcomes and survey quality indicators in this type of data collection.

First, we can investigate the interviewer effects on the discrepancies between reported participation in the Supplemental Nutrition Assistance Program (SNAP; formerly known as the Food Stamp Program) in FoodAPS, which is an important measure in FoodAPS, and actual SNAP participation. With the high rates of consent for record linkage to SNAP administrative files (i.e. 97·5%), we can utilize most of the FoodAPS data for this investigation. Past literature studying the efficacy of food stamps on health and well-being has often cited possible misreporting as a critical challenge[13,14].

We can also consider interviewer effects on measures of food security and income, as these variables have strong associations with food acquisition[15,16]. Food security and income are generally determined based on sets of sensitive questions. The USDA defines household food security as 'access by all people at all times to enough food for an active, healthy life'[17] and FoodAPS used a standard set of ten questions to estimate household food security. FoodAPS also used a complex set of questions designed to capture many sources of income at the individual level. Given their complexity and sensitivity, these two measures may be vulnerable to the influence of the interviewers.

In terms of survey quality, in the present paper we also investigate interviewer effects on the time between the initial (pre-diary) and final (post-diary) face-to-face interviews, henceforth termed the 'interview day gap'. Although there is more than one way to ensure better survey quality for a survey like FoodAPS, careful adherence to the interview timeline with a household is very important, as household circumstances such as monthly income might change from month to month. Adherence to the interview timeline would also provide data users with the best temporal associations between the measures collected in different components of the survey (i.e. the household screening interview, the initial interview, the diary surveys and the final interview). Thus far, interviewer effects on interview timing have not been studied in the food acquisition survey context, and this is an important first step in understanding how interviewers contribute to the variance in time taken to complete the several parts of a food acquisition study[18].

In summary, the present study seeks to understand the effects of FoodAPS interviewers on the accuracy of SNAP reports, the delay in interview timing, item-missing values in income measures, and the variance in income and food security measures. We aim to examine interviewer effects on these survey outcomes after accounting for differences that are attributable to geographical, household and interviewer characteristics. The magnitude of these effects may indicate that there is a non-ignorable correlation between the responses obtained by an interviewer that has to be corrected for in substantive analyses involving the data set. From the survey operations perspective, this is also indicative of the survey data quality.

Methods

FoodAPS

FoodAPS was a 7 d diary survey of food acquisition that collected data from a nationally representative, stratified multistage sample of non-institutionalized households in the continental USA with a total sample of 14 317 individuals living in 4826 households. FoodAPS targeted four strata: (i) SNAP households; (ii) non-SNAP households with income below the federal poverty guideline; (iii) non-SNAP households with income ≥100–185 % of the poverty guideline; and (iv) non-SNAP households with

income ≥185 % of the poverty guideline at the time of data collection (April 2012–January 2013). Households' survey eligibility was determined based on SNAP participation, household size and household income. The overall weighted study response rate was 41·5 % (RR3 calculation of the American Association for Public Opinion Research)[19]. FoodAPS was approved by the institutional review board of the survey contractor (Mathematica Policy Research).

Study protocol

Prior to the diary survey period, the primary food planner in a sampled household was identified as the primary respondent (PR). The PR was interviewed in two in-person interviews at the beginning (i.e. initial interview) and the end of the diary survey (i.e. final interview). The interviewer trained the PR on the usage of the food book and a scanner was provided at the initial interview for tracking bar-coded foods during the week. The PR was then asked to train other household members to track their own food acquisitions. The training was supplemented with a video depicting how to fill in the food book and how to use the food scanner. Interviewers were instructed to run through a practice grocery trip scenario with the PR at the end of the video to ensure that the instructions were understood. The PR also was asked to call a centralized location three times during the week to report household food events. A telephone interviewer would call them if they did not call in as scheduled. Our study specifically focused on the interviewers who worked with the sampled households on a face-to-face basis.

Figure 1 provides an overview of the FoodAPS survey protocol, while Table 1 provides an overview of the type of data collected at each interview.

Interviewers

Nearly 200 interviewers were hired to conduct the face-to-face interviews. The interviewers were provided with both an at-home training (where they had to track their own food acquisition for a week) and on-site training which lasted four or five days and covered: general training skills for inexperienced interviewers; specific training for screening households and conducting the initial and final interviews; and how to train respondents to use the food logs. At the end of the training, interviewers were evaluated on their knowledge and performance, and they were certified for data collection. Interviewers were assigned to geographical locations closest to their home. Every region was overseen by a field manager who monitored interviewer productivity in terms of response rates and contact rates. These data were not available for analysis.

Given the importance of average interviewer workload in determining the expected impacts of variable interviewer effects on the variance of survey estimates[5], Table 2 summarizes the total number of interviewers (n) and their average workloads (i.e. counts of households or persons providing a value for that outcome) for each of the outcomes investigated in the present study. In Table 2, the average total workload for measures of food security, income and missing income are from the final interview, while SNAP discrepancy is from the initial interview. Interview day gap is based on the dates of the initial and final interviews, being constrained to households that had the same interviewer for both interviews. A total of 165 interviewers conducted the initial and final interviews for the same household. The workload reflects the number of initial and final interview pairs conducted on the same household. The individual income and missing income workloads are based on individual respondents rather
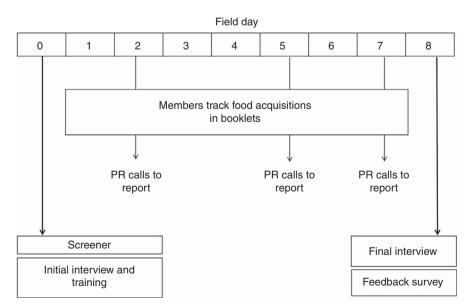

Fig. 1 Overview of the planned data collection week of FoodAPS (PR, primary respondent; FoodAPS, National Household Food Acquisition and Purchase Survey)

Table 1 Summary of information collected at each interview in FoodAPS

Initial interview	Telephone interview	Final interview
HH information: • HH size • Relationship of HH members to PR • Sociodemographic information of each HH member Participation in food programmes: • SNAP status • Free lunch for children • Food programme information Food purchasing habits: • Most common location to get food • Access to other food sources • No. of times each HH member eats out	Food at-home and food-away-from-home events by HH member	HH members' eating habits: • Frequency of preparing dinner • Frequency of eating together • Frequency of having guests over Diet: • HH members' perception of the healthiness of their diet • HH members' dietary habits • Food insecurity HH members' sources of income HH expenses: • Housing • Transportation • Utilities • Insurance • Medical • Child/adult care Major life events Financial practices: • Frequency of reviewing bills • Punctuality of bill payment • Difficulty with bill payment

FoodAPS, National Household Food Acquisition and Purchase Survey; HH, household; PR, primary respondent; SNAP, Supplemental Nutrition Assistance Program.

Table 2 Counts of interviewers and average workloads by outcome in FoodAPS

Outcome	n	Average interviewer workload
Food security (final interview)	169	28·56 households
SNAP discrepancy (initial interview)	163	24·44 households
Interview day gap (both interviews)†	165	28·40 households
Individual income (final interview)	169	84·72 individuals
Missing income (final interview)	169	84·72 individuals

FoodAPS, National Household Food Acquisition and Purchase Survey; SNAP, Supplemental Nutrition Assistance Program.
†Measured only for households in which the same interviewer conducted both the initial and final interviews.

than households, as the PR was asked about the income of each person aged 16 years or older within the household.

Outcome variables

As mentioned previously, five different survey outcomes were examined in the present study. Each outcome is described below, after restricting the sample to only those who have complete records on all the covariates of interest (detailed in the next section).

Food insecurity

Food insecurity, measured from the ten-item, 30 d USDA Food Security Module for adults, was collected at the final interview at the household level[20]. The sum of the scores for the scale ranges from 0 (no food insecurity) to 10 (highest food insecurity). In the present study, scores of 0–2 and 3–10 were collapsed into two separate categories, yielding a binary food security measure 'food security' v. 'food insecurity', as per USDA definitions[17].

SNAP discrepancy

SNAP discrepancy was measured by matching household-level self-reports of current SNAP status collected at the initial interview against administrative records of their SNAP status. Prior consent to match their responses was obtained in the FoodAPS interview. Observations were coded as 'discrepant' when their reported SNAP status did not match administrative records.

Interview day gap

Although the planned data collection period spanned eight days, this was not always the case for a given household. The range of the data collection period spanned from 6 to 167 d, with an average of 9·8 d and a median of 8·0 d. Due to the skewed distribution, with a high number of interviews finishing on time (i.e. within 6–8 d), we grouped the raw measure into three ordinal categories (on time, 6–8 d; delayed, 9–10 d; extremely delayed, >10 d).

Total individual income

An individual's monthly income was constructed from separate questions in the final interview about six different types of income for each household member who was at least 16 years old. In the final FoodAPS data set, any missing income at the individual level was imputed. As the interest in the present study was in interviewer effects on the measurement variance of this outcome, total individual income prior to imputation was used, and the analysis was restricted to individuals who reported any income at all, leaving 6753 observations for analysis.

Missing income

This variable was constructed from the total individual income variable. Observations that were flagged as having

any imputed income amount (i.e. any item-missing on the six types of income) were coded as missing income, while all other observations were coded as non-missing.

Covariates

A total of eighteen household-level covariates, eight individual-level covariates and three interviewer-level covariates (see Tables 3 and 4) that might be associated with the five outcomes were identified from the FoodAPS data set and Census 2010 data. Covariates that were used in outcome variable construction, such as current SNAP status for the SNAP discrepancy outcome variable, were not included in the respective models. Food insecurity was dropped as a covariate for the food insecurity model. Employment status of the PR was dropped for the income

and missing income models. In the model for income, household income was dropped as well. In the model for interview day gap, where the sample was restricted to households who were interviewed by the same interviewer at the initial and final interview, the interviewer familiarity covariate was dropped. Individual covariates were included only in the income and the missing income models.

Statistical analysis

To examine interviewer effects on the five outcomes of interest, we fit separate generalized linear mixed models to each of the outcomes using PROC GLIMMIX in the statistical software package SAS version 9.4. We used a cumulative logit link for the interview day gap outcome

Table 3 Household-level/individual-level covariates from FoodAPS included in the models

Variable group	Covariate	Type	Reason for inclusion
Socio-economic status	• Employment status of PR (employed, unemployed)†	Categorical	Socio-economic status of the household is associated with income, food security, and potentially interview day gap
	• Household income†,‡	Continuous	
	• Job status change (past 3 months) (yes, no)†	Categorical	
	• Car ownership/lease (owns/rents, does not own/rent)†	Categorical	
	• HH size†,‡	Continuous	
	• SNAP status (SNAP, non-SNAP)†	Categorical	
	• Food security rating (0 = food security, 1 = food insecurity)†	Categorical	
Geographic information	• Location (rural, non-rural)	Categorical	Included to control for the differential assignments of the interviewers to locations with very different profiles[6]. Poverty ratio was calculated by dividing the total population count under 100 % of the poverty level by the total population of the block group
	• Block group population density (divided into four quartiles)	Categorical	
	• Poverty ratio	Continuous	
	• Number of restaurants within 5 mile (8 km) radius†	Continuous	
Survey experience	• Language of interview (English, non-English)†	Categorical	Included to control for the perceived difficulty of the survey
	• Interviewer familiarity (different interviewer, same interviewer)	Categorical	
	• Feedback form: freq. of completing food book (every day, …, did not complete at all, missing)	Categorical	
	• Feedback form: difficulty for HH members (very easy, …, very difficult, not applicable, missing)	Categorical	
	• Feedback form: ease of keeping track of food (very easy, …, very difficult, missing)	Categorical	
Cooperation	• Inbound calls (no. of calls dialled to the telephone centre from HH)	Continuous	Included to control for the respondent's cooperativeness with the survey
	• Outbound calls (no. of calls dialled from the telephone centre to HH)	Continuous	
	• Completion of income worksheet (yes, no)†	Categorical	
Individual covariates	• Age†	Continuous	Included to control for the respondents' characteristics, and their relationship to the PR
	• Sex (male, female)†	Categorical	
	• Race (White, Black, others, multiple race)	Categorical	
	• Marital status (married, widowed, divorced, separated, never married)	Categorical	
	• Education level (below HS, HS, college or above)	Categorical	
	• Financial condition (1 = very comfortable and secure, 5 = in over your head)	Categorical	
	• Frequency of paying bills on time (never, rarely, sometimes, usually, always, missing)	Categorical	
	• Frequency of reviewing bills for accuracy (never, rarely, sometimes, usually, always, missing)	Categorical	

FoodAPS, National Household Food Acquisition and Purchase Survey; PR, primary respondent; HH, household; SNAP, Supplementary Nutrition Assistance Program; freq., frequency; HS, high school.
†Tested as random slopes, all not significant at $\alpha = 0.05$.
‡Income was rescaled by 10 000; income and household size were mean-centred (i.e. the overall mean of the variable was subtracted from each individual value).

Table 4 Interviewer-level covariates from FoodAPS included in the models

Variable group	Covariate	Type	Reasons for inclusion
Interviewer experience	• No. of interviews completed†	Continuous	These variables approximate interviewer[4] experience, which could be an explanation for interviewer effects on these outcomes
	• Ever interviewed in metropolitan locations (non-metropolitan only, metropolitan only, both metropolitan and non-metropolitan)	Categorical	
	• Travelled across PSU (yes, no)	Categorical	

FoodAPS, National Household Food Acquisition and Purchase Survey; PSU, primary sampling unit.
†Number of interviews was mean-centred.

(modelling the probability of a longer day gap), and a binary logit link for food insecurity (predicting the event of food insecurity), SNAP discrepancy (predicting the event of having a discrepancy) and missing income outcomes (predicting the event of missing income). We assumed that the income outcome followed a marginal normal distribution and employed an identity link for that model.

For each of these outcomes, we fit an unconditional model first, with a fixed overall intercept, a random interviewer effect (assumed to be normally distributed with a mean of 0 and constant variance) and a residual (for income, normally distributed with a mean of 0 and constant variance). In the individual models, a random household effect was added as well to account for the clustering within households. Our objective in fitting the initial unconditional model was to obtain an estimate of the raw between-interviewer variance in the random interviewer effects, prior to adjustment for the fixed effects of any covariates.

Following that, we fit a model including fixed effects of the household-level covariates to examine the change in interviewer variance after accounting for the possible household-level correlates of the outcomes. If a given interviewer worked on very similar households, these covariates may behave like interviewer-level covariates (having minimal variation within interviewers) and their subsequent fixed effects may then explain interviewer variance. Finally, we added the fixed effects of interviewer-level covariates to the model, again to examine the change in the interviewer variance after accounting for possible interviewer-level correlates of the outcomes. Because all models used the exact same cases with non-missing values on all variables (given the restrictions noted above), the interviewer variance was consistently estimated using the same cases across all steps. The variance components were tested for significant differences from zero using likelihood ratio tests based on mixtures of χ^2 probabilities[21].

The food insecurity and interview day gap models were estimated using the adaptive Gaussian quadrature (AGQ) method with ten integration points, while the individual income model was estimated using Laplace approximation, and the SNAP discrepancy and missing income models were estimated using the less computationally intensive (and potentially more biased) residual pseudo-likelihood estimation technique[11]. Given the relatively rare occurrence of SNAP discrepancies and the computational difficulty in fitting both a household and interviewer random intercept on a binary outcome, this approach was mainly used to get an approximate sense of any interviewer variance; the small prevalence prevented the AGQ from converging and producing a valid solution.

The possibility that the effects of particular covariates varied across interviewers was also tested in a purely exploratory fashion for all models (i.e. we did not have any theoretical expectations about particular relationships of the covariates with the outcomes that may vary across interviewers). The covariates for which we tested the possibility of random slopes are indicated in Table 2. The estimates given herein are all unweighted.

For each of these models, observations that were missing on any of the covariates were dropped from analysis. At each model-fitting step, an intraclass correlation coefficient (ICC) was calculated to determine how much of the total variance in a given outcome was accounted for by the interviewers. A high ICC would indicate that the interviewers were contributing substantially to the overall variance; the values observed in typical survey practice range between 0·01 (very small) to 0·12 or more (very large)[12]. The ICC was calculated with the following formula:

$$ICC = \frac{\sigma_0^2}{\sigma_0^2 + \sigma^2},$$

where σ_0^2 is the variance of the random interviewer effects and σ^2 is the unexplained residual variance within interviewers (fixed to $\pi^2/3$, or the fixed variance of the logistic distribution, for the logistic models). In the case of the income models, the ICC was calculated with the following formula, where the only addition is the variance of the random household effect denoted by $\sigma_{0,\mathrm{HH}}^2$:

$$ICC = \frac{\sigma_0^2}{\sigma_0^2 + \sigma_{0,\mathrm{HH}}^2 + \sigma^2}.$$

After fitting the model that included the fixed effects of the household- (or individual-) and interviewer-level covariates, empirical best linear unbiased predictors (EBLUP) of the random interviewer effects in the model were computed and examined graphically. We first identified potential outliers in the distribution of the predicted random interviewer effects. We then profiled those interviewers found to be highly unusual and who could be influencing our estimates of the interviewer variance.

Finally, the interviewer variance components were estimated again after excluding these potential outliers. Although there are other methods to identify outliers, such as using the random effects variance shift outlier model (RVSOM), which was developed for linear mixed models and is computationally more intensive, we opted for a graphical method as our outcomes are mostly categorical[22].

The multiplicative interviewer effect on the variance of a descriptive estimate[5] was then calculated from the ICC after removal of the outliers, using the following formula:

$$\text{Interviewer effect} = 1 + (\overline{b} - 1) \times \text{ICC},$$

where $\overline{b}$ is the average interviewer workload for that outcome. For example, if the average interviewer workload $\overline{b}$ is thirty interviews and the ICC is 0·10, there will be an almost 300 % increase in variance of a descriptive estimate (e.g. an estimated proportion) due to interviewer effects.

Results

Descriptive statistics

For those households with complete data on the covariates of interest, 72·15 % reported having 'food security', while the remaining 27·85 % reported having 'food insecurity'. Out of the 94·05 % of households that provided match consent, only 3·48 % of the households had a discrepant report of their current SNAP status. As for interview day gap, 57·9 % of the final interviews were completed on time, 24·1 % were delayed and the remaining 18·1 % were extremely delayed. The range of total individual monthly income (in US dollars) was $4·17 to more than $90 000, with

a median of $1660·00, a mean of $2329·34 and sd of $2846·55. Due to the skewed distribution of the responses, individual income was log-transformed for analysis. Of the 10 182 individuals who were 16 years or older, 13·83 % had missing income data.

Interviewer effects

Due to very small cell sizes which led to the food insecurity model not converging, location of household, employment status of PR and household income were dropped as covariates in this model. Table 5 provides a summary of the estimated variances of the random interviewer effects in each of the models considered for the household-level outcomes (i.e. food insecurity, SNAP discrepancy, interview day gap), while Table 6 provides estimated variance components for the individual-level outcomes (i.e. income and missing income). For the outcomes, there is evidence of significant variation in the random interviewer effects in the unconditional model (Model 1), with the variability accounted for by the interviewers ranging from 8·30 % (SNAP discrepancy in Table 5) to 11·50 % (interview day gap in Table 5).

The addition of fixed effects of the household-level covariates (i.e. Model 2) reduced the variance of the random interviewer effects for each of the five outcomes, as expected, but this reduction varied substantially in magnitude across the outcomes. For food insecurity and log-income, the variance of the random interviewer effects dropped by 54·46 % (equating to a change in the ICC from 0·0836 to 0·0399) and 83·02 % (ICC = 0·0727 to 0·0177), respectively, while for the interview day gap and missing income outcomes the reduction was less pronounced. The estimated variance dropped by 3·39 % for interviewer

Table 5 Interviewer effects on the household-level outcomes in FoodAPS, with standard errors

	Food insecurity		SNAP discrepancy		Interview day gap	
	(variable type = binary)		(variable type = binary)		(variable type = ordinal)	
	Full sample (n 4808)	Outlier removed (n 4758)	Full sample (n 3650)	Outlier removed –	Full sample (n 3513)	Outlier removed (n 3488)
Variance of random interviewer effects						
Model 1	0·3002***	–	0·2985*	–	0·4278***	–
SE	0·0629	–	0·1642	–	0·0863	–
ICC	0·0836	–	0·0830	–	0·1150	–
Model 2	0·1367***	–	0·1852	–	0·4133***	–
SE	0·0434	–	0·1557	–	0·0857	–
ICC	0·0399	–	<0·0001	–	0·1116	–
Model 3	0·1143***	0·0921**	0·1961	–	0·3690***	0·3297***
SE	0·0410	0·0379	0·1608	–	0·0781	0·0725
ICC	0·0336	0·0272	<0·0001	–	0·1008	0·0911
Notes	AGQ (10 points)		RSPL		AGQ (10 points)	

FoodAPS, National Household Food Acquisition and Purchase Survey; SNAP, Supplementary Nutrition Assistance Program; ICC, intraclass correlation coefficient (for ordinal and binary outcomes, calculated using ICC = $\sigma_0^2 / (\sigma_0^2 + 3·29)$); AGQ, adaptive Gaussian quadrature; RSPL, restricted pseudo likelihood, Newton–Raphson optimization; HH, household.
Model 1, unconditional model; Model 2, model with HH covariates; Model 3, model with HH and interviewer covariates; Model 3 outlier removed, model with HH and interviewer covariates (outlier interviewer removed).
Tested with mixture χ^2 test, first fitting the model with random household intercept, then the model with interviewer and household random intercepts; none of the models tested had evidence of random slopes. Rescaling of the interviewer variance component across the binary logit and cumulative logit models, given the household- and individual-level covariates included in Model 2, did not yield substantially different findings[24].
*P < 0·05, **P < 0·01, ***P < 0·0001.

Table 6 Interviewer effects on the individual-level outcomes in FoodAPS, with standard errors

	Log-income		Missing income	
	(variable type = continuous)		(variable type = binary)	
	Full sample (n 6755)	Outlier removed (n 6657)	Full sample (n 10 182)	Outlier removed (n 10 130)
Variance of random intercept				
Model 1				
Interviewer	0·0577***	–	0·3773***	–
SE	0·0097	–	0·0853	–
HH	0·0996***	–	2·3290***	–
SE	0·0132	–	0·1192	–
Interviewer ICC	0·0727	–	0·0629	–
Model 2				
Interviewer	0·0098***	–	0·2524***	–
SE	0·0028	–	0·0782	–
HH	0·0349***	–	2·3812***	–
SE	0·0092	–	0·0782	–
Interviewer ICC	0·0177	–	0·0426	–
Model 3				
Interviewer	0·0084***	0·0040*	0·2477***	0·2363***
SE	0·0027	0·0022	0·0769	0·0737
HH	0·0345***	0·0328***	2·3870	2·3943***
SE	0·0092	0·0093	0·1262	0·1270
Interviewer ICC	0·0153	0·0072	0·0418	0·0399
Notes	Laplace		RSPL	

FoodAPS, National Household Food Acquisition and Purchase Survey; HH, household; ICC, intraclass correlation coefficient (for binary outcome, calculated using ICC $= \sigma_0^2 / (\sigma_0^2 + \sigma_{HH}^2 + 3 \cdot 29)$); RSPL, restricted pseudo likelihood, Newton–Raphson optimization.
Model 1, unconditional model; Model 2, model with HH covariates; Model 3, model with HH and interviewer covariates; Model 3 outlier removed, model with HH and interviewer covariates (outlier interviewer removed).
Tested with mixture χ^2 test, first fitting the model with random household intercept, then the model with interviewer and household random intercepts; none of the models tested had evidence of random slopes. Rescaling of the interviewer variance component across the binary logit and cumulative logit models, given the household- and individual-level covariates included in Model 2, did not yield substantially different findings[24].
*$P < 0 \cdot 05$, **$P < 0 \cdot 01$, ***$P < 0 \cdot 0001$.

day gap and 33·10 % for missing income. For the SNAP discrepancy model, adding the household-level covariates dropped the variance by 37·96 % (ICC = 0·0830 to <0·0001) and the variance component was no longer significantly different from zero. Despite these reductions due to the addition of the household-level (and individual-level) covariates, there was still a significant amount of unexplained variance among the interviewers for four of the five models.

The addition of the fixed effects of the interviewer-level covariates (i.e. Model 3) did not change the variance of the random interviewer effects substantially in the missing income model; the variances of the random interviewer effects dropped by 1·86 %. The decreases in the variances of the random interviewer effects for the food insecurity model, interview day gap model and the income model were more substantial (i.e. 16·39 % in the food insecurity model, 10·72 % in the interview day gap model and 14·29 % in the log-income model).

Overall, the biggest drop in the variance for most of the models occurred after the inclusion of the fixed effects of the household-level covariates, suggesting that household-level features and the types of households recruited by the interviewers were playing a larger role in introducing variance in the outcomes among interviewers compared with interviewer-level covariates. However, neither the household-level covariates nor the interviewer-level covariates did well in reducing the variance of the random interviewer effects for the interview day gap model.

Examination of the empirical best linear unbiased predictors to identify outliers

Examination of the EBLUP of the random interviewer effects indicated that there were outliers in each of these models, except for SNAP discrepancy. From the four Q–Q plots of the EBLUP estimates (Fig. 2), we visually identified the most severe outlier for each outcome (circled on the plots). For illustration purposes, only one outlier has been selected for examination from the food insecurity model as it is less clear compared with the rest which is the most severe. We then examined these outliers based on covariates that are influential to the outcome (refer to Tables 7 and 8; covariates suppressed in the tables can be found in the online supplementary material, Supplemental Tables 1 to 4).

When the single outlier in the food insecurity model (Fig. 2(a)) was examined, we found that the households worked by this interviewer reported less food insecurity than expected (more seem to be food secure). Language of interview was a significant predictor of food insecurity, with non-English interviews being more likely to report food insecurity (OR = 1·4408, 95 % CI 1·0948, 1·8961; Table 7). This interviewer conducted non-English
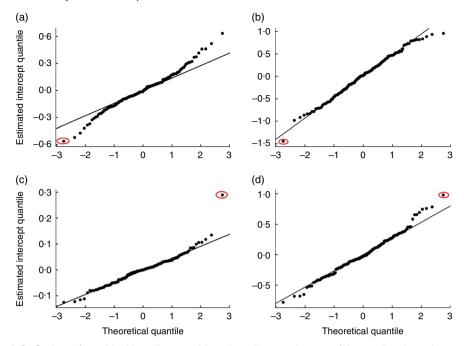
Fig. 2 (colour online) Q–Q plots of empirical best linear unbiased predictor estimates of the random interviewer effects in FoodAPS for: (a) food insecurity; (b) interview day gap; (c) income; (d) missing income. The most severe outlier for each outcome is circled (FoodAPS, National Household Food Acquisition and Purchase Survey)

Table 7 Odds ratios and 95 % confidence intervals for models with household and interviewer covariates (household models)†

	Food insecurity		SNAP discrepancy		Interview day gap	
	OR	95 % CI	OR	95 % CI	OR	95 % CI
Household covariates						
Non-English interview	1·4408	1·0948, 1·8961				
Non-SNAP household	0·3493	0·3011, 0·4052				
Does not own/rent car	1·6193	1·3434, 1·9517				
Poverty ratio	2·8098	1·7319, 4·5581	4·8778	1·4915, 15·9539		
Household income			0·2547	0·0949, 0·6839		
Household size			1·1803	1·0627, 1·3110		
PR not employed					0·8196	0·7070, 0·9500
Rural					0·7392	0·5951, 0·9182
Inbound calls					0·8516	0·8083, 0·8973
Interviewer covariates						
Did not travel across PSU	0·7285	0·5761, 0·9213				
Number of interviews					0·9945	0·9913, 0·9976

SNAP, Supplementary Nutrition Assistance Program; PR, primary respondent; primary sampling unit.
†Only covariates significant at $P < 0.01$ are shown. Significance was tested with Wald test based on all parameters associated with the covariate.

interviews in twenty-one households but only one of them reported food insecurity, which was unusual. Compared with the whole sample, about 40 % of households with non-English interviews reported food insecurity.

The interviewer identified for the interview day gap model (Fig. 2(b)) was unusual because this interviewer did better than expected. This interviewer completed the final interview within the 8 d timeline for all twenty-five households assigned, despite half of the interviewed households having an employed PR. Based on the interview day gap model, PR who are not employed are less likely to take longer (OR = 0·8196, 95 % CI 0·7070, 0·9500; which means PR who are employed are more likely to take longer).

The outlier interviewer for the log-income model (Fig. 2(c)) was unusual in that total incomes reported by most individuals in this interviewer's households were roughly double their predicted income values based on the model. This interviewer was working primarily in rural areas with low poverty (100 % poverty rate of 0·005 to about 0·088 across the block groups worked). As can be seen in Fig. 3, three other interviewers who interviewed in areas with the same profile (Fig. 3(b), 3(c) and 3(d)) have a lower discrepancy between actual and predicted income. When the data from the outlier interviewer (Fig. 3(a)) were examined more closely, most respondents were recorded as receiving their reported earnings twice per month, which were then doubled during data processing

Table 8 Estimates/odds ratios and 95 % confidence intervals for models with household and interviewer covariates (individual models)†

	Log-income		Missing income	
	β	95 % CI	OR	95 % CI
Household covariates				
Non-SNAP household	0·3181	0·2715, 0·3647		
Does not own/rent car	−0·1982	−0·2599, −0·1365		
Food secure	0·0827	0·0324, 0·1330		
Poverty ratio	−0·5246	−0·6654, −0·3838		
Household income			0·5721	0·4275, 0·7657
Did not complete income worksheet			2·2662	1·8808, 2·7305
Individual covariates				
Male	0·2746	0·2388, 0·3104		
Marital status (Ref. = never married)				
Married	0·3516	0·3005, 0·4027		
Widowed	0·2917	0·1970, 0·3864		
Divorced	0·3663	0·3036, 0·4290		
Separated	0·2728	0·1745, 0·3711		
Education level (Ref. = college or above)				
Below HS	−0·5593	−0·6186, −0·5000		
HS	−0·3369	−0·3796, −0·2942		
Frequency review bills (Ref. = always)				
Never	0·0975	0·0309, 0·1641		
Rarely	0·0533	−0·0246, 0·1311		
Sometimes	0·0199	−0·0343, 0·0742		
Usually	0·0980	0·0454, 0·1505		
Missing	0·0291	−0·2396, 0·2978		
Financial condition (Ref. = in over your head)				
Very comfortable and secure	0·5830	0·4749, 0·6911		
Able to make ends meet without much difficulty	0·3906	0·2914, 0·4898		
Occasional difficulty making ends meet	0·2600	0·1653, 0·3547		
Tough to make ends meet but keeping head above water	0·1540	0·0610, 0·2470		

SNAP, Supplementary Nutrition Assistance Program; Ref., reference category; HS, high school.
†Only covariates significant at $P < 0.01$ are shown. Significance was tested with Wald test based on all parameters associated with the covariate.
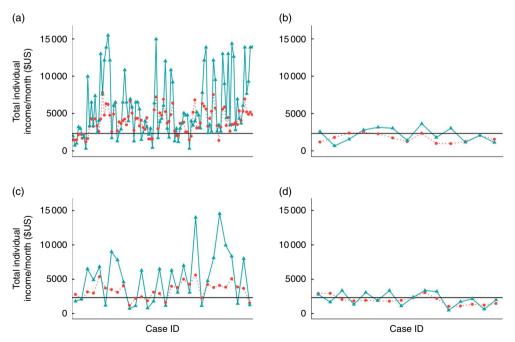

Fig. 3 (colour online) Predicted (—●—) *v.* actual (—▲—) individual income by interviewer in FoodAPS: (a) outlier: interviewer 490; (b) interviewer 469; (c) interviewer 531; (d) interviewer 663 (FoodAPS, National Household Food Acquisition and Purchase Survey)

to estimate the earnings portion of total monthly income. This pattern deviated from the majority of the sample, which indicated receiving their reported earnings once per month.

As for the missing income model (Fig. 2(d)), this interviewer was unusual because the degree of missingness for the income measure was higher than expected. In the full

sample in this model, of those who completed the income worksheet prior to the final interview, only 9·50 % had any missing income measures at all. However, among those interviewed by this interviewer who had completed the income worksheet, 38·00 % of them had at least one missing income measure.

Even after removal of these outliers, a significant amount of interviewer variance remained (Tables 5 and 6). The income model saw a substantial change in the ICC (from ICC = 0·0153 to 0·0072), or a 52·38 % reduction of the variance component of the interviewer after the removal of the outlier. The change was less dramatic in the other three models, with a 19·42 % reduction in the food insecurity model, a 10·65 % reduction in the interview day gap model and a 4·60 % reduction in the missing income model.

Multiplicative interviewer effects on variances of descriptive estimates

The ICC computed after the removal of the outliers were used in calculating the multiplicative interviewer effects for descriptive estimates (i.e. proportion indicating any food insecurity and mean individual income). This was not done for interview day gap and missing income, as these were constructed variables that measure survey quality. For food insecurity (unweighted mean = 0·2806, $\mathrm{SE_{SRS}} = 0·0065$; descriptive statistics estimated after removal of outliers and assuming simple random sampling), where the average interviewer workload was $\bar{b} = 28·56$ (Table 1) and the ICC was estimated to be 0·0272 (without the outlier), the resulting multiplicative interviewer effect was 1·75. Computing the standard error that accounts for the variance inflation due to the interviewer effects gives: $\mathrm{SE_{INT}} = \sqrt{(0·0065^2 \times 1·75)} = 0·0086$.

Similarly, for individual income (unweighted mean = 2277·38, $\mathrm{SE_{SRS}} = 34·19$), the interviewer effect was calculated to be 1·60 and, using the same equation as above, the standard error accounting for interviewer effects was 43·25 (a roughly 26 % increase). To illustrate the improvement in the estimates by removing the outlier, the estimated standard error accounting for interviewer effects prior to the removal of outliers for food insecurity is 0·0090 (interviewer effect = 1·93, unweighted mean = 0·2785, $\mathrm{SE_{SRS}} = 0·0065$). The standard error estimate of 0·0086 after removal of the outliers is only a minor improvement (4·44 %) over 0·0090.

It should be noted that although the ICC ranged from 0·0072 (very small) to 0·0911 (large) after the removal of the outlier interviewers, overall multiplicative interviewer effects on the variances of estimates are a function of both the ICC and interviewer workload. With an ICC of 0·0072, if the average interviewer workload is 100 instead, the interviewer effect would have been 1·70 instead of 1·60 and the standard error accounting for interviewer effects would have increased by 31 %. Therefore, even a small ICC may be cause for concern, depending on the average interviewer workload.

Discussion

The current study presented evidence of significant unexplained variance among FoodAPS interviewers in terms of reported values for interview day gaps and presence of missing income values, even after accounting for a variety of household- and interviewer-level covariates. Levels of unexplained variance among interviewers for measures of any food insecurity and total income were much lower. We conclude that some interviewer effects are indeed present in FoodAPS, and this raises concerns about the need to deal with interviewer effects in food acquisition surveys more generally for both data users and data producers[15,16]. Different interviewers were somehow eliciting different responses (or responses of varying quality) from respondents despite efforts to standardize their training. A failure of the secondary analyst to account for these interviewer effects in their substantive models may lead to overly liberal conclusions (i.e. increased type I error rate) and understated standard errors of key estimates as interviewer effects would inflate standard errors. It is also possible that different interviewers were successfully recruiting and interviewing different types of respondents, which could lead to different responses being elicited as well[7,8,12,23]. As there was a lack of data on interviewer characteristics available from the FoodAPS, it is not possible with the current data set to determine if interviewers were in fact recruiting respondents more similar to themselves, as has been found in previous research[5,12].

In practice, this problem could be reduced by having more standardization in the interviewing protocol, more rigorous interviewer training and, most importantly, field monitoring of large deviations from the patterns of the data collected. In the case of the possibility that interviewers were recruiting different types of respondents, efforts could be made to monitor the profiles of respondents recruited by each interviewer and intervene if the respondent profiles lack diversity. However, monitoring interviewer-level deviations from the pattern of responses collected may be slightly challenging; using multilevel modelling for such 'live' monitoring has not been done during survey operations before, to the authors' knowledge. The effectiveness of this type of monitoring for interviewer effects during a data collection period requires future research.

In the analysis presented above, the current paper attempted to identify interviewers who may be contributing heavily to the interviewer effects. We were able to identify outliers for four of the models investigated. Removal of interviewers that were found to be outliers helped to reduce the variance for the individual income model, but not substantially for the other models. We also profiled these interviewers and tried to determine why their data were unusual.

This work therefore provides preliminary evidence of the efficacy of using multilevel modelling techniques to

monitor interviewers in the field during data collection. By being able to monitor in real time, errors in data collection could be reduced. This approach does need some empirical backing before it could be implemented, so this monitoring could be started a few weeks into the data collection period or after some agreed-upon data collection threshold. RVSOM could also be implemented to identify influential outliers where possible to make outlier monitoring less subjective to human judgement[22].

Limitations

The main limitation for many studies on interviewer effects is the lack of randomization of interviewers to respondents, and this is true for the present study as well. To keep costs low, FoodAPS interviewers tended to be assigned to a few locations that were near their homes to reduce travel time and costs. The infeasibility of such a design means that interviewer effects may be confounded with geographical effects[7]. The current paper tries to address this problem by including geographic information as covariates, but there could be geographic information that was not accounted for. Given this limitation, if important geographic covariates have been omitted from the models (which simply were not available in the data set), the effects of these covariates on the outcomes would be attributed to the interviewers. If a given interviewer also tended to work households with very similar characteristics, those household characteristics could also be driving the unexplained interviewer variance. We made a distinct effort to account for as many relevant household covariates as we could in the models fitted (Table 2).

The current data set also lacks information on the interviewers themselves, which is another major limitation; interviewer characteristics could be a factor in the responses they receive[5]. It would have been ideal to be able to include interviewer characteristics such as interviewer experience, race, age and gender, but the FoodAPS contractor did not provide any information about its interviewers. The present study attempted to include variables that could indicate interviewer experience (e.g. greater caseload could indicate a more experienced interviewer; refer to Table 4 for proxy covariates included). As with the geographical information, these models could have been improved by having more information on the interviewers.

More generally, we analysed only five outcome measures that we deemed likely to be prone to interviewer effects and/or important for a wide variety of users of FoodAPS data (e.g. any food insecurity). There may be interviewer effects on other FoodAPS measures as well.

Conclusion

Despite these limitations, the fact that a multilevel modelling approach was able to identify unusual interviewers

lends credibility to our suggested approach of 'live' field monitoring. With a well-specified model for important outcomes and survey quality measures, this approach could potentially identify interviewers who may be performing differently, leading to recommendations for retraining. Ultimately, this could reduce interviewer effects for the outcomes that are important to a given survey, which in the case of FoodAPS would be any food insecurity and other variables related to food acquisition.

Acknowledgements

Supplementary material

To view supplementary material for this article, please visit https://doi.org/10.1017/S1368980018000137

References

1. Ricciuto LE & Tarasuk VS (2007) An examination of income-related disparities in the nutritional quality of food selections among Canadian households from 1986–2001. *Soc Sci Med* **64**, 186–198.
2. Marchioni DM, Claro RM, Levy RB *et al.* (2011) Patterns of food acquisition in Brazilian households and associated factors: a population-based survey. *Public Health Nutr* **14**, 1586–1592.
3. Griffin MK & Sobal J (2014) Consumer food system participation: a community analysis. *Ecol Food Nutr* **53**, 579–595.
4. US Department of Agriculture, Economic Research Service (2017) National Household Food Acquisition and Purchase Survey (FoodAPS). http://www.ers.usda.gov/foodaps (accessed October 2017).
5. West BT & Blom AG (2017) Explaining interviewer effects: a research synthesis. *J Surv Stat Methodol* **5**, 175–211.
6. Eckman S & Kreuter F (2013) Undercoverage rates and undercoverage bias in traditional housing unit listing. *Sociol Methods Res* **42**, 264–293.
7. O'Muircheartaigh C & Campanelli P (1999) A multilevel exploration of the role of interviewers in survey non-response. *J R Stat Soc Ser A Stat Soc* **162**, 437–446.
8. Korbmacher J & Schroeder M (2013) Consent when linking survey data with administrative records: the role of the interviewer. *Surv Res Methods* **7**, 115–131.

9. Schaeffer NC, Dykema J & Maynard DW (2010) Interviewers and interviewing. In *Handbook of Survey Research*, 2nd ed., pp. 437–470 [PV Marsden and JD Wright, editors] Bingley: Emerald Group Publishing Limited.

10. Liu M & Stainback K (2013) Interviewer gender effects on survey responses to marriage-related questions. *Public Opin Q* **77**, 606–618.

11. Kim Y, Choi Y & Emery S (2013) Logistic regression with multiple random effects: a simulation study of estimation methods and statistical packages. *Am Stat* **67**, 171–182.

12. West BT & Olson K (2010) How much of interviewer variance is really non-response error variance? *Public Opin Q* **74**, 1004–1026.

13. Kreider B, Pepper JV, Gundersen C et al. (2012) Identifying the effects of SNAP (Food Stamps) on child health outcomes when participation is endogenous and misreported. *J Am Stat Assoc* **107**, 958–975.

14. Gundersen C, Kreider B & Pepper J (2012) The impact of the National School Lunch Program on child health: a nonparametric bounds analysis. *J Econom* **166**, 79–91.

15. Turrell G, Hewitt B, Patterson C et al. (2002) Socioeconomic differences in food purchasing behaviour and suggested implications for diet-related health promotion. *J Hum Nutr Diet* **15**, 355–364.

16. Turrell G & Kavanagh AM (2006) Socio-economic pathways to diet: modelling the association between socio-economic position and food purchasing behaviour. *Public Health Nutr* **9**, 375–383.

17. US Department of Agriculture (2016) Definitions of Food Security. http://www.ers.usda.gov/topics/food-nutrition-assistance/food-security-in-the-us/definitions-of-food-security/ (accessed April 2017).

18. Hu M, Gremel GW, Kirlin JA et al. (2007) Nonresponse and underreporting errors increase over the data collection week based on paradata from the national household food acquisition and purchase survey. *J Nutr* **147**, 964–975.

19. American Association for Public Opinion Research (2015) Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys. http://www.aapor.org/AAPOR_Main/media/publications/Standard-Definitions2015_8thedition withchanges_April2015_logo.pdf (accessed February 2017).

20. US Department of Agriculture (2012) US Adult Food Security Survey Module: Three-Stage Design, With Screeners. http://www.ers.usda.gov/media/8279/ad2012.pdf (accessed November 2017).

21. West BT, Welch KB & Galecki AT (2014) *Linear Mixed Models: A Practical Guide Using Statistical Software*, 3rd ed. Boca Raton: FL: Chapman Hall/CRC Press.

22. Gumedze FN & Jackson D (2011) A random effects variance shift model for detecting and accommodating outliers in meta-analysis. *BMC Med Res Methodol* **11**, 19.

23. West BT, Kreuter F & Jaenichen U (2013) 'Interviewer' effects in face-to-face surveys: a function of sampling, measurement error or nonresponse? *J Off Stat* **29**, 277–297.

24. Hox JJ, Moerbeek M and van de Schoot R (2018) *Multilevel Analysis: Techniques and Applications*, 3rd ed. New York/ Abingdon: Routledge.