Journal of Official Statistics, Vol. 26, No. 1, 2010, pp. 165–191

Get It or Drop It? Cost-Benefit Analysis of Attempts to Interview in Household Surveys

Dmitri Romanov[1] *and Michal Nir*[1]

We develop a cost-benefit model for streamlining allocation of field staff efforts invested in attempting to interview the designated sample units. By accounting for heterogeneous response propensity among various population groups, the costs of fieldwork, and the utility of the information gathered in a survey, the model provides a guideline for determining the optimal maximum number of attempts to interview. Limiting interview attempts lowers the survey's response rate, possibly creating a nonresponse bias – a factor that is not directly reflected in a cost-benefit reckoning. We demonstrate the use of the model by simulating different limits on the number of attempts to conduct a face-to-face interview in Israel's Household Expenditure Survey. Under the most stringent simulated limit – three attempts to interview in the Arab sector and four attempts in the Jewish sector, in which the response rate falls from 88% to 76% – we found that limiting the number of attempts to interview causes no significant bias in estimates of the main survey variables.

Key words: Fieldwork efficiency; nonresponse bias; limit on attempts to interview; duration analysis model.

1. Introduction

While the survey practitioners may prefer to get response from all sampled units, budget constraints that apply to any given survey make it necessary to forego nonrespondents, at some point. Moreover, the survey managers' decisions regarding where and when to desist from further attempts to interview "hard" cases are usually arbitrary and ad hoc relative to the circumstances of the case at hand.

It is a well-established fact that extended interviewer efforts invested in contacting hard-to-get respondents and in converting refusals tend to reduce the nonresponse bias (e.g., Stoop 2004; Lynn et al. 2002). This is not to say that low response rates necessarily lead to high nonresponse bias (Abraham et al. 2006). Moreover, raising the response rate does not assure survey quality. As Sturgis et al. (2006) put it, "The use of arbitrary "large numbers" as response rate benchmark criteria is of dubious utility for assessing survey quality. It is perfectly possible for a survey meeting such criteria to produce more biased estimates than a different survey, which fails to meet the criteria. More emphasis should be placed, where possible, on bias assessments which compare responders to nonresponders on key survey variables" (p. 10). Groves (2006) concluded in a comprehensive study, "there is little

[1] Israel's Central Bureau of Statistics, 66 Kanfei Nesharim, corner Bachi str., Jerusalem 95464, Israel.
Emails: dromanov@cbs.gov.il and michaln@cbs.gov.il
Acknowledgments: We are grateful to Liat Nokrian, Tzahi Makovky and Arie Reiter from the ICBS for assistance. We are also indebted to the Editor and to the referees for their instructive remarks and suggestions.

empirical support for the notion that low response rate surveys de facto produce estimates with high nonresponse bias" (p. 670).

Furthermore, several studies have shown that the higher the desired response rate, the higher the marginal cost of interviewing (Lynn and Clarke 2002; Teitler et al. 2003; Philippens et al. 2004). It follows that the practical utility of a higher response rate in terms of survey quality may be quite limited, whereas every increase in the response rate inflicts an escalating cost on the surveying organization.

The literature on survey methodology rarely deals with the economic aspects of fieldwork organization. Purdon et al. (1999) calculated the cost of specific alternatives for the performance of the Family Resources survey but ruled out the model's generalization because of the survey design complexity. In the most detailed analysis of survey costs to date, Groves (2004) focused on the relationship between survey errors, costs and sample design, while discussing a variety of structural cost models and providing general guidelines of cost-conscious fieldwork organization. When aiming to derive a design feature that minimizes an error source subject to a given cost constraint, Groves barely addressed the *dual problem* that bothers probably every statistical organization: what are the ways and means of saving money on a survey with given design, without magnifying the survey errors.

The present study asks whether it is possible to analytically devise an optimal limit on the number of attempts to interview, as an across-the-board rule for a set of comparable surveys, while minimizing the possible bias of survey estimates. In this context, "attempts to interview" mean any effort made by an interviewer in purpose of conducting the interview. Hence, we do not distinguish between attempting to contact the subject, i.e., dealing with noncontacts, and attempting to convert refusals.

We advance the literature in two directions. First, we develop a cost-benefit model that allows a survey manager to coherently devise the limit on the number of attempts to interview, by accounting for key survey characteristics: length of interview, response rate and fieldwork costs. The model is applicable to a wide variety of the most expensive surveys – which include personal or telephone interviews. The model may be used also for establishing, within one survey, different rules for population groups that are expected to exhibit different response rates. Second, we introduce a duration analysis model – which is routinely used in biometrics and econometrics – for examining the factors associated with probability of success in interviewing the survey subjects, conditional on past (vain) attempts to interview.

We demonstrate the use of the cost-benefit model by simulating two different limits on the number of attempts to interview in the first phase of the Household Expenditure Survey carried out by Israel's Central Bureau of Statistics.

The rest of this study is organized as follows. Section 2 presents the cost–benefit model of the interviewing process, from which we derive a rule for determining an optimal limit on the number of attempts to interview. In Section 3 we examine the effects of the characteristics of interviewer and household on the number of attempts to interview, in the first phase of Israel's Household Expenditure Survey, by estimating a duration analysis model. Section 4 presents a simulation of two limits on the number of attempts to interview and compares the survey estimates from the actual sample to those that were received from the truncated samples with lower response rates. This comparison serves to gauge any nonresponse bias that may have been created by imposing the limit on the number of attempts to interview. The final section discusses issues related to the application of the proposed model in household surveys' fieldwork.

2. The Cost-Benefit Model

In the following schematic model we consider the survey costs and benefits as a function of attempts to interview. The model applies to surveys of which the target population is either households (usually sampled from a dwellings register) or individuals (usually sampled from a population register). In this general setup, the interviewer has to visit the home of the sampled household/individual in order to interview. If he/she fails to interview the subject (either noncontact or refusal) in the first attempt, further attempts are made by additional visits of the interviewer.

The surveying organization has to decide about the number of the attempts to interview. Let $n_{ij} = 1, \ldots, N_{ij}$ denote the number of attempts until interviewer j interviews subject i. We assume that the quality of each attempt is constant for all interviewers in a given survey, who all make their best effort to interview the designated sample.

The final result of the attempts to interview is either a success (a complete interview), $R_{ij} = 1$, or a failure (nonresponse), $R_{ij} = 0$. For simplicity, let us assume that only the number of attempts to interview affects the response rate. The effects of interviewers' and respondents' observable and nonobservable characteristics on the likelihood of interviewing are discussed in Section 3.

The marginal social cost function of the attempts to interview is denoted as $c(n_{ij})$. The marginal cost includes all variable costs, such as travel and locating time (during the first interview attempt), transport and interview time, and a respective share of all relevant fixed costs, such as supervision. The above-mentioned *administrative* costs are only a part of the *social* cost of conducting a survey by a national statistical organization. The social cost includes also the subject's response burden, which is affected, *inter alia*, by interview length, frequency of contacts with subject, and his/her interest in the survey topic. Against the response burden one should offset the utility of the information received by the subject as a result of the survey. For example, participation in a household expenditure survey gives to the respondent an accurate picture of expenses and income. (For a general model of marginal efficiency cost of public funds, see Slemrod and Yitzhaki 2002).

Let us define the social utility of interviewing subject i as the social value of the information that the subject is asked to provide. We denote this value by B_i. Since participation in mandatory surveys is a sort of a "lump-sum tax" – which is imposed on randomly selected subjects – the utility is equal for every subject, i.e., $B_i = B$ for all i, irrespective of the number of attempts to interview that is needed. The social utility is realized only if the survey questionnaire is completed, hence the expected utility is defined as:

$$\mathrm{E}(B_i) = B \cdot \Pr(R_{ij} = 1) \tag{1}$$

Generally, the probability of interviewing depends on the number of attempts to interview:

$$\Pr(R_{ij} = 1) = F(n_{ij}) \tag{2}$$

The marginal probability of interviewing, $f(n_{ij}) \equiv dF/dn_{ij}$, is a decreasing function of the number of attempts to interview, i.e., $f'(n_{ij}) < 0$. Therefore, the marginal expected utility declines with the number of attempts to interview: $d\mathrm{E}'(B)/dn_{ij} < 0$.

Thus, by expressing the social utility in monetary terms and juxtaposing it to the social cost of interviewing, we may determine an optimum number of attempts to interview. In Figure 1, this number is noted at point $n*$ (indices of n_{ij} are omitted for notational simplicity): every attempt to interview up to this point will deliver, in expectation, positive social utility net of cost; every attempt to interview beyond this point will waste social resources. For national statistical organizations, funded by taxpayers' money, a socially optimal number of attempts to interview turns out to be the optimal maximum number of attempts.

It should be noted that social cost and utility functions are actually stepwise functions, because n takes only discreet values. Therefore, the respective marginal functions are discontinuous. Both curves are presented in Figure 1 as continuous for simplicity sake, just to illustrate the optimization idea.

Figure 1 reflects a situation in which the probability and the expected utility of interviewing are uniform for the entire sample. In practice, this does not happen due to heterogeneity in response propensity among different groups of survey subjects, associated with a variety of cultural, demographic and socio-economic attributes. Accordingly, Figure 2 shows, as example, two marginal expected utility curves. Type $\underline{B}$ subjects are "easy" to interview: it takes only few attempts to reach them; higher order attempts to interview them are mostly fruitless – the response rate falls steeply with increasing n. In contrast, Type $\overline{B}$ subjects require, in expectation, more attempts to interview, as reflected in a thicker and longer right-hand tail of $f(n_{ij})$. Accordingly, these traits should be taken into account in determining the optimum number of attempts to interview, $\underline{n}^* < \overline{n}^*$. Otherwise, setting for both groups a single standard optimum number of attempts to interview would result in excessive attempts to interview Type $\underline{B}$ subjects and failure to fulfill the response potential of Type $\overline{B}$ subjects.

The technique for applying the model is discussed below.

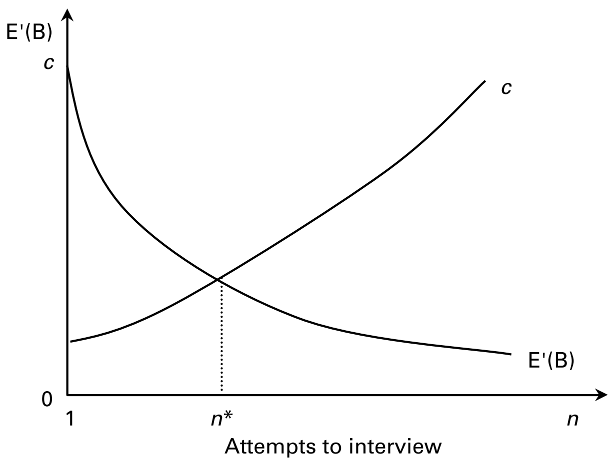

Fig. 1. Marginal Cost, Marginal Expected Benefit, and Optimum Number of Attempts to Interview

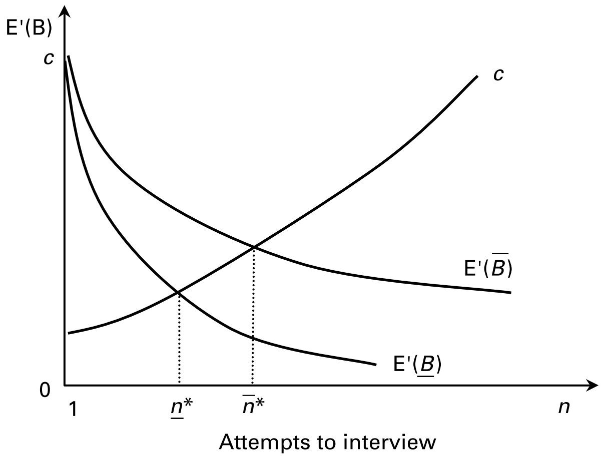

Fig. 2. Marginal Cost, Marginal Expected Benefit, and Optimum Number of Attempts to Interview, for Two Populations Differentiated in Response Propensity

2.1. Model Application

To apply the model, three kinds of estimates are needed:

(1) An estimate of the marginal cost function of the attempts to interview, $c(n_{ij})$;
(2) An estimate of the marginal probability of interviewing function, $f(n_{ij})$;
(3) A monetary estimate of social utility of interviewing the subject, B.

The **marginal cost function of the attempts to interview** in the survey depends on how the fieldwork is organized, but in any case the accounting system of the statistical organization should provide more or less detailed estimates of this function.

The **marginal probability of interviewing function** is estimated, for each survey and in accordance with the characteristics of the households in the survey, by duration analysis models (e.g., see Kiefer 1988, for a survey, and Greene 2000, for a general presentation). If N is the number of attempts needed to interview a given subject (subject's index omitted for notational simplicity), then the survival function is defined, in terms of the number of attempts to interview, as the probability of noninterview before attempt n:

$$S(n) \equiv \Pr(N > n) = 1 - F(n) \tag{3}$$

In the next section, we estimate the survival function by means of nonparametric analysis (the Kaplan-Meier estimator). We also estimate a parametric model of the accelerated failure time type, assuming a linear effect of the covariates on the response variable, which is the number of attempts until accomplishing the subject's interview:

$$\Pr(\mathrm{N} > n|\mathbf{x_c}) = \Pr(\mathrm{N}_0 > \exp(-\mathbf{x_c}'\beta)n) \tag{4}$$

where $\mathbf{x_c}$ is an array of covariates; N_0 is an event, $R_{ij} = 1$, sampled from the baseline distribution corresponding to values of zero for the covariates. The relevant distribution is chosen by examining the correspondence between the data and several theoretical distributions; we chose the exponential distribution specification. Array $\mathbf{x_c}$ can be partitioned into variables that are known before and after the interview. The first group includes the interviewer characteristics and the characteristics of the subject's residential neighborhood. In Israel, for example, the subject's neighborhood usually attests to his/her ethnic group, for the vast majority of Jews and Arabs live in segregated neighborhoods, and even to the degree of religiosity, in Jewish neighborhoods. The second group includes all variables reported by the subject in the interview; obviously, these are known only for the respondents who completed the interview.

The **monetary estimate of the social utility of interviewing the subject** cannot be calculated. As a proxy, one may suggest the time that the subjects need to complete the survey questionnaire (average interview length, T), being evaluated as the average value of leisure, θ_T, under the assumption that the interviews are conducted during the leisure hours. That is, we would like to estimate $\hat{B} = T \cdot \theta_T$. Unfortunately, there is no consensus among economists regarding the valuation of leisure time that the respondent forfeits when participating in a survey. Furthermore, even if leisure time could be evaluated somehow, it would be an underestimate of the *social utility* of the information elicited by the survey, because this information is a public good that exerts positive externality effects.

Lacking a credible monetary estimate of the social utility, we cannot arrive at an *absolute* determination of the optimum number of attempts to interview in a given survey. However, we may approximate the social utility of interview in survey k in the form of $B_k = T_k \theta_T \mu$, where $\mu > 1$ corrects for the positive externality effects of survey information. Under the reasonable assumption that θ_T and μ are constants for all surveys, we may construct a social-utility ratio between a pair of comparable surveys: $B_k / B_m = T_k / T_m$. The transformation that expresses the relative social utility of two surveys as a ratio of the interview lengths rests on the assumption that the production of surveys by statistical organizations is efficient, at margin, in terms of the response burden. This assumption seems to be reasonable in view of systematic efforts invested by the national statistical organizations in survey planning and management – from selecting the issues for investigation, through developing and testing the questionnaire, to administering the fieldwork – in order to minimize the response burden. This is not to say, however, that the survey's burden is directly and completely reflected by the questionnaire's length. Survey literature discusses other factors influencing survey's burden such as topic interest, survey's sponsors and incentives for participating (e.g., Groves et al. 2004; Groves et al. 2006). The effects of these factors are shown to differ according to the attributes of the sampled person and survey design features (e.g., Groves et al. 2000).

So, we use the interview length – that can be objectively measured and compared between the surveys – as a proxy for *social utility*. All other factors influencing the willingness of subjects to participate in the survey and the feeling of burden of respondents – which are evidently on the *social cost* side of the model – are implicitly reflected in the hazard rate, $h(n)$:

$$h(n) = f(n) / S(n) \tag{5}$$

From the definition of the survival function (Formula 3) and the fact that $f(n)$ is a decreasing function of the number of attempts to interview follows that the more burdensome is a survey the lower is its hazard rate.

As discussed above, the survey social cost includes two factors: administrative costs and response burden. Hence, the bottom line of our cost-benefit analysis – the difference between social utility and social cost – is a function of three variables: interview length, T, marginal cost of the attempts to interview, $c(n)$, and hazard rate, $h(n)$.

$$Benefit - Cost = \varphi(T, c(n), h(n)) \tag{6}$$

where $d\varphi/dT > 0$, $d\varphi/dc < 0$, $d\varphi/dh < 0$.

Now, to make the model operational, one has to choose one survey as a point of reference and for this survey to determine – arbitrarily, by rules of thumb, organizational tradition, or by trial and error – the maximum number of attempts to interview, n_0^*. Then, assuming a multiplicative form of φ, we can derive, for any other comparable survey k, the maximum number of attempts, n_k^*,:

$$n_k^* = n_0^* \times \frac{T_k / \left(c_k\left(n_0^*\right) h_k\left(n_0^*\right)\right)}{T_0 / \left(c_0\left(n_0^*\right) h_0\left(n_0^*\right)\right)} = n_0^* \times \frac{T_k}{T_0} \times \frac{h_0\left(n_0^*\right)}{h_k\left(n_0^*\right)} \times \frac{c_0\left(n_0^*\right)}{c_k\left(n_0^*\right)} \tag{7}$$

Formula (7) states that the maximum number of attempts to interview in survey k will be greater when completing its questionnaire is more time-consuming, its hazard rate is lower, and its marginal fieldwork costs are smaller – all relative to respective parameters of the reference survey. Note that the marginal cost and the hazard rate of survey k are evaluated at the point n_0^*.

In this manner, by a pairwise comparison, it is possible to derive consistently and uniformly the limits on the number of attempts to interview for all comparable surveys performed by a statistical organization.

As to the question which two surveys may be considered comparable, in order to streamline the number of interview attempts across them, we posit that at very least they should be similar in the target population and in the sampling units.

Formula (7) may be applied to various subpopulations in one survey, with one subpopulation serving as a reference point. In this case, questionnaire length and marginal fieldwork cost will be approximately equal for the different subpopulations, since the content of the questionnaire and the collection method are the same. Thus, the maximum number of attempts to interview will be determined solely by the ratio of the hazard rates, as in our simulation (Section 4).

It should be emphasized that the cost-benefit model does not account for nonresponse bias, which may increase (as a result of or arise if it was not present before) imposing a limit on the number of interview attempts. This issue has to be handled, consequently, by simulation, as done for example in Olson 2006. In Section 4 we apply our model on Israel's Household Expenditure survey and check for the appearance of the nonresponse bias.

3. Duration Analysis of the Probability of Interviewing

This section empirically investigates on the one hand the relationship between the characteristics of locality, interviewer and subjects, and on the other hand the number of interview attempts that are needed to complete Part A of Israel's 2004 Household

Expenditure survey (henceforth, the HES) questionnaire. Before presenting our findings we cursorily describe the survey's fieldwork process.

The HES is a mandatory survey, performed in three phases. In Phase 1, the interviewer sets out to locate the sampled household. The number of attempts to interview in this phase includes attempts to locate the household, to make contact with an eligible member of the household, and to interview the latter using Part A of the questionnaire, which solicits demographic details about all members of the household. Also in this phase, the interviewer introduces an expenditure diary, which members of the household need to fill in during a two-week period. In Phase 2, which lasts about two weeks, the members of the household are visited regularly (at least four times). On each visit, the interviewer examines the record of expenditures in the diary and, if needed, helps the household members to fill in correctly the figures required, and encourages the household members to continue recording all their expenditures. In Phase 3, the interviewer makes the final visit and completes a detailed account of large and/or irregular outlays for a variety of goods and services that were not covered in the diary. Also during this visit, Part B of the questionnaire is used to interview about the household's durable goods, income and work specifics. As a rule, all three phases of the HES are performed by the same interviewer. These practices are believed to be crucial for establishing good working relations with the members of the household. The result is a response rate of more than 88% in a survey that is considered one of the most difficult to perform in view of the onerous burden that it imposes on the respondents. Interviewing in the Household Expenditure survey takes place during the period 16.00–21.00, Sunday (a working day in Israel) to Thursday, unless upon the first attempt the subject asks to be visited in the morning or early afternoon hours.

We focused on Phase 1 of the survey because it resembles a typical survey using the face-to-face interview, in which visits to the sampled households have the aim of contacting eligible members and persuading them to be interviewed.

Table 1 shows significant differences between Jews and Arabs in the probability of successful interviewing in the HES, with a higher response rate in the Arab sector. Also, relatively more Arabs than Jews were interviewed upon the interviewer's first attempt (53% vs 34%, respectively). Shinar et al. (2005) suggest that such differences reflect a unique cultural characteristic of the Arab population – the custom of hospitality, which becomes prominent since the interviewers in the Arab sector are also Arab. Locality size does not seem to affect the probability of interviewing. The locality's socioeconomic level also appears to have no perceptible effect on the response rate. This is contrary to findings in other countries, where the differences in response rates across the socioeconomic and urban-rural lines were found (e.g., Stoop 2004; Feskens et al. 2007).

We test the effects of these factors below by means of a multivariate regression. The interviewer's personal characteristics, professional experience and skills are known to affect the duration and the outcome of the interviewing process (Groves and Couper 1998; Lynn et al. 2002; Hox and de Leeuw 2002). Table 2 shows the demographic traits of the interviewers who worked on the 2004 Household Expenditure survey. We have no data as to the behaviors and the techniques interviewers apply to contact and persuade subjects to participate in the survey. Insofar as these factors affect interviewer

Table 1. Response Rates by Sector, Locality Size and Locality Socioeconomic Level[a] (percent)

	Total	Jews	Arabs
Total sample	100.0	100.0	100.0
Noneligible	10.8	10.7	11.8
Refusals	4.2	4.4	2.1
Noncontacts	4.8	5.2	1.1
Partial completion of questionnaire	1.1	1.0	1.9
Complete interview (excl. noneligible cases)	88.6	88.0	94.2
Thereof: on first attempt	36.2	34.4	53.0
Response rate, out of eligible sample, by locality size			
Population up to 2,000	<90.3>	<90.3>	
2,000–50,000	90.1	88.8	
50,000–100,000	89.5	88.5	95.0
100,000–200,000	87.0	87.0	98.3
Jerusalem	79.0	79.0	
Haifa	91.0	91.2	<85.7>
Tel Aviv-Yafo	93.0	93.1	<90.5>
Response rate, out of eligible sample, by socioeconomic index deciles			
1–2	93.9	91.1	96.0
3–4	85.8	83.5	93.7
5–6	87.4	87.4	
7–8	90.3	90.4	88.6
9–10	87.6	87.6	

Source: 2004 Israel's Household Expenditure Survey, authors' computations.
[a] "Jews" includes non-Arabs; "Arabs" includes Moslems, Christians, and Druze. Locality size is as of 2004. Socioeconomic level is based on the ICBS 2001 socioeconomic index, which has a scale of 1 (low) to 10 (high). < > = less than 20 observations in cell.

effectiveness, a statistical model would account for their influence as an interviewer's fixed effect, which would presumably reflect their "trade secrets," motivation and other unobserved traits.

Most interviewers who work on the HES are Jewish since Jews account for most of the sample and because, due to language requirements and cultural differences, Arab interviewers work mainly with Arab and Druze subjects. As a result we cannot determine whether the higher response rate among Arabs is attributable to the characteristics of the Arab sector or to the superior performance of Arab interviewers, relative to Jewish interviewers.

Most interviewers are women. More than 70% are married; more than one-third have academic degrees. Arab interviewers are better educated and younger than Jewish interviewers; they also have less seniority as interviewers with Israel's Central Bureau of Statistics (the ICBS).

Figure 3 depicts the results of the nonparametric estimation of the marginal probability of interviewing using the Kaplan-Meier method. The figure shows clearly that both sectors evince a similar marginal probability of response for attempts two to seven. Thus, the discrepancy in the final response rates of Jews and Arabs traces to a higher response rate among Arabs upon the first attempt, the effect of which is partly offset by more attempts to interview (beyond seven) in the Jewish sector.

Table 2. Response Rates and Characteristics of Interviewers in Household Expenditure Survey, by Sector[a] (percent unless otherwise noted)

	All Interviewers	Jews	Arabs
Number of interviewers	66	59	7
Complete interview (excl. noneligible cases)	88.9	88.2	95.8
Interviewer characteristics			
Women	81.8	89.8	14.3
Immigrants (after 1989)	22.7	25.4	
Marital status	100.0	100.0	100.0
Single	7.6	5.1	28.6
Married	71.2	71.2	71.4
Divorced	15.2	17.0	
Widowed	6.0	6.7	
Education	100.0	100.0	100.0
Secondary	62.1	66.1	28.6
Bachelor's degree	16.7	13.6	42.8
Master's degree	21.2	20.3	28.6
Average age, years	49.6	51.1	37.2
Average seniority as interviewer, weeks	165.7	167.8	147.9

Source: 2004 Israel's Household Expenditure Survey, authors' computations.
[a] 206 observations omitted because their records lacked interviewer details.

By performing a multivariate analysis of the cumulative probability of interviewing the subject, we examine the combined impact of the characteristics of locality, interviewer, and subject. Our model estimates the effect of these traits on the probability of obtaining the interview conditional on the number of attempts, by means of the survival function estimation (Formula 4). This model is quite different from the two statistical models used in empirical research concerning nonresponse.

In one popular model, researchers estimate a binary response (logit or probit) regression that focuses on the effects of various characteristics on the probability of completing the interview of a subject in a given attempt (e.g., Groves and Couper 1998; Stoop 2004).

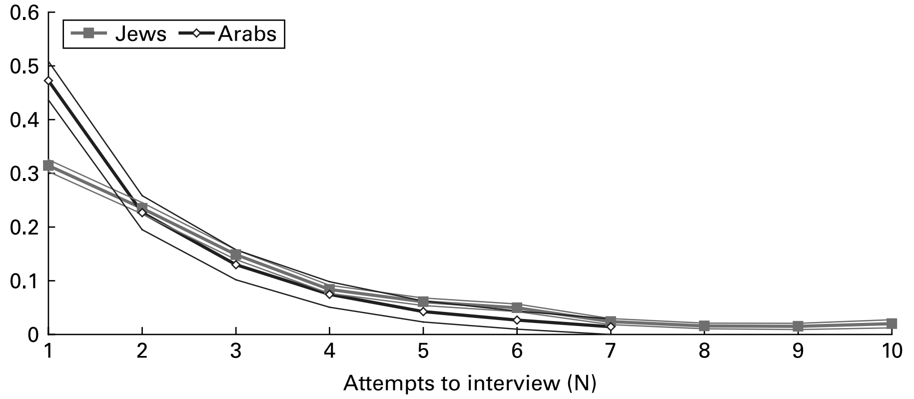

Fig. 3. Marginal Probability of Interviewing, by Number of Attempts, by Sector (with 95% confidence interval). Source: 2004 Israel's Household Expenditure Survey, authors' computations

In fact, a binary regression "slices up" the continuity of the interviewing process, first looking at the conditional distribution $R_{ij}|n_{ij} = 1$, then analyzing conditional distribution $R_{ij}|n_{ij} = 2$, and so on, till $R_{ij}|n_{ij} = N$. If the estimates of the regression coefficients change systematically as the number of attempts to interview rises, a researcher concludes that the populations interviewed in high-order attempts are dissimilar to those who responded in the first attempts. It should be noted that statistical inference concerning the differences between the estimates of two slice-wise regressions seems to be problematic.

Nicoletti and Peracchi (2005), who tested for dependency between the probability of contacting the subject and the probability of his/her cooperation, estimated a censored bivariate probit model. In their model, the number of attempts to interview served as an explanatory variable in the likelihood-of-contact equation, and the sign of this variable, as expected, was negative (statistically significant).

Unlike our statistical model, both of these estimation methods are susceptible to problems stemming from the truncation of the distribution of results, if the interviewing process is terminated before all cases are dealt with and their final status clarified. Furthermore, both methods disregard the information contained in the serial number of each attempt to interview, i.e., they do not exploit the fact that the marginal probability of interviewing declines with the number of attempts.

Table 3 presents the results of the survival function estimation in four specifications. Model (a) includes only locality characteristics; Model (b) adds interviewer characteristics; Model (c) adds sixty dummy variables representing the interviewer fixed effects, and Model (d), without interviewer fixed effects, includes observed characteristics of the subjects from the completed interviews. Sixty-three observations of subjects who were visited by six interviewers whose workload contained fewer than twenty subjects were deleted from the estimation. The estimation was performed without an intercept. Though it is usual in the literature to account for the time of visitation (e.g., Purdon et al. 1999; Philippens et al. 2004), we do not include this factor among the explanatory variables, because in the HES, as a rule, interviews take place in the evening.

The explained variable is the survival rate, i.e., the probability that the subject *will not be interviewed* in a given number of attempts. Therefore, a positive sign of an estimate indicates a variable that increases the number of attempts to interview – one that correlates negatively with the probability of interviewing the subject upon the first attempt.

The estimates of Model (a) indicate that more attempts are needed to interview Jews than to interview Arabs. Also, the households sampled from the three largest cities in Israel (Jerusalem, Haifa, and Tel Aviv-Yafo) are more difficult to interview than those in smaller localities. When interviewer characteristics are added, in Model (b), the effect of subject's sector (Jews versus Arabs) loses its statistical significance, and the significance of the dummy variable for Jerusalem also disappears. As for interviewer characteristics, gender, age, and seniority were not found to have any effect on the number of attempts to interview. This last finding is somewhat surprising, especially since ongoing training is applied for all interviewers working at Israel's Central Bureau of Statistics, who are required periodically to attend workshops and seminars directed to improving their professional skills. One might expect duration of employment to reflect an accumulation of skill that would allow interviewers to do better in persuading reluctant subjects (Groves and Couper 1998; Hox and de Leeuw 2002) as well as to contact "hard to get" subjects

Table 3. Estimates of Survival Function of Attempts to Interview[a]

Independent variable	(a)	(b)	(c)	(d)
Jewish locality[b]	0.384 (0.037)*	−0.103 (0.092)	−0.158 (0.060)*	
Jerusalem	0.081 (0.042)	0.061 (0.047)	0.038 (0.093)	−0.111 (0.044)*
Haifa	0.138 (0.042)*	0.150 (0.045)*	0.280 (0.082)*	0.131 (0.040)*
Tel Aviv-Yafo	0.215 (0.039)*	0.187 (0.043)*	0.216 (0.097)*	0.064 (0.039)**
Locality socioeconomic level[c]	−0.001 (0.007)	−0.012 (0.007)	0.000 (0.011)	−0.012 (0.007)**
Interviewer characteristics				
Bachelor's degree (control group: secondary education)		−0.047 (0.033)	−2.939 (0.421)*	0.016 (0.030)
Master's degree		−0.118 (0.027)*	−0.958 (0.301)*	−0.091 (0.024)*
Marital status, unmarried (control group: married)		−0.043 (0.025)**	0.016 (0.209)	−0.032 (0.023)
Women		−0.007 (0.038)	−1.023 (0.226)*	0.002 (0.035)
Jewish		0.576 (0.102)*		0.218 (0.072)*
Age		−0.000 (0.001)	0.074 (0.010)*	0.004 (0.001)*
Seniority with ICBS as interviewer, years		−0.001 (0.004)	0.079 (0.024)*	−0.007 (0.004)**
Household characteristics				
Owns dwelling				0.027 (0.022)
Jewish head of household				0.016 (0.054)
Age of head of household				−0.004 (0.001)*
Years of schooling of head of household				0.002 (0.002)
Self-employed head of household				0.042 (0.033)
Non-working head of household				−0.080 (0.027)*
Persons in household				−0.032 (0.006)*
Log of household per-capita expenditure				0.011 (0.015)
Number of breadwinners				−0.052 (0.014)*
Subject and interviewer from same town				0.129 (0.022)*
Interviewer's fixed effect	None	None	Yes[d]	None
Observations (*N*)	7,576	7,411	7,348	5,850
Log-likelihood	−8,865.2	−8,656.9	−8,442.3	−6,232.0

Source: 2004 Israel's Household Expenditure Survey, authors' computations.

[a] Standard deviations are in parentheses. * 0.05 significant estimate, ** 0.1 significant estimate. Omitted 206 observations for which interviewer details were not recorded.

[b] The variable is defined according to the religion of a majority of its inhabitants by the 1995 Census. "Jews" includes non-Arabs; "Arabs" includes Moslems, Christians, and Druze.

[c] Socioeconomic level is based on the 2001 socioeconomic index, on a scale of 1 (low) to 10 (high).

[d] Of 60 fixed effects, 40 were significant at the 0.05 level.

(O'Muircheartaigh and Campanelli 1999; Durrant and Steele 2007). This finding may possibly be explained by the fact that senior interviewers are assigned to more difficult areas or have a larger workload: an average annual workload of an interviewer with fewer than two years' seniority in the HES counts 95 households, as against 127 for interviewers with more than two years' seniority. These two factors may offset the effect of the superior skill of a veteran interviewer as against a newly hired interviewer who has not yet accumulated specific experience in fieldwork. It was also found in Model (b) that interviewer academic education is associated with a lower number of attempts (to a statistically significant extent among holders of Master's degrees). Unmarried interviewers are not more effective than married ones; Jewish interviewers have to make more attempts than Arab interviewers to reach for the subjects in their workload.

The insertion of the interviewer fixed effects in Model (c) makes it impossible to pinpoint the effect of the interviewers' sector. Therefore, the effect of the Jewish localities becomes significant, as in Model (a), although at much less intensity. Once the fixed effects are included, the effect of interviewers' education on reducing the number of attempts becomes stronger. Interviewer age and seniority with ICBS were also found to correlate positively with the number of attempts, and women interviewers were found to be more effective than male interviewers. Importantly, the socioeconomic level of the locality has not been found to have any statistically significant effect in any model so far. It should be mentioned that usually interviewers in the HES receive a monthly "portion" of addresses to cover – in geographical proximity one to other – for the sake of fieldwork efficiency. For that reason, the interviewer fixed effect may induce correlation in the outcomes for the households in close addresses covered by the same interviewer, although the dispersion of monthly portions around the city and between the cities mitigates this problem to a large extent.

Model (d) adds characteristics of head of household (the individual whom the household members place at the top of the list) and other household characteristics. The demographic and occupational characteristics of the subjects in the Household Expenditure survey come into sight only after the subjects complete Part A of the questionnaire; their levels of expenditure and income become clear only if they complete the expenditure diary and Part B of the questionnaire. Accordingly, in Model (d), households that did not belong to the survey population, those that did not respond, and those that did not complete all parts of the survey questionnaire as required were deleted from the sample, which clearly lessens the degree of comparability between Model (d) and the first three models. Model (d) is estimated on the sample of 5,850 households, with the average number of interview attempts 2.33, against 2.58 in the sample used in the Models (a)–(c).

The results of Model (d) show that when one controls for the effects of major cities subjects' characteristics, the socioeconomic level of the locality correlates negatively with the number of attempts to interview, meaning that interviewing is easier in wealthier localities. Fewer attempts are needed to interview households that are larger and have older and nonworking heads (who are presumably easier to find at home). Controlling for other factors, the level of per-capita household expenditure was not found to have a separate effect, but the effect of the household economic status may be reflected in the negative sign of the number-of-breadwinners variable. These findings are consistent with the results of nonparametric multiple regression based on the extended Gini index by Schechtman et al. (2008), who investigated Israel's Household Expenditure Surveys in 1997–2001 and found that the response rate is a positive function of the household's income and size. They found also that the Arab population tends to response more than the Jewish one.

The effect of interviewer characteristics is consistent with the findings discussed above: interviewers who hold Master's degrees are more effective than their less-educated peers and Arab interviewers require fewer attempts to interview than Jewish ones, controlled for subject's religion. The signs of the age and seniority variables indicate that interviewers' performance diminishes with age, *ceteris paribus*. However, interviewers with greater professional experience manage to interview their subjects in fewer attempts. When subject and interviewer live in the same town, more attempts to interview are made. To explain this, one may assume that interviewers tend to carry out more visits than are required in places that are near their homes and along their routes of work.

4. Simulation of Limiting the Number of Interview Attempts

In this section we perform two simulations in which the number of attempts to interview is limited differentially for subjects in Jewish and Arab sectors. The limit would be imposed nonrandomly for those subjects who have not been interviewed when the limit is reached. Accordingly, the distribution of interview attempts would be truncated at that limit.

The Arab sector will serve as a reference point in which we set, arbitrarily, two alternative limits: three and four attempts. On the basis of Formula (7), the maximum number of attempts in the Jewish sector is set in accordance with the ratio of the hazard rates in the two sectors that were estimated in the previous section (Figure 3):

Attempt number	1	2	3	4	5
Arab subjects	0.6188	0.5478	0.5509	0.5586	0.5714
Jewish subjects	0.3729	0.4133	0.3944	0.3225	0.3245
Ratio of hazard rates	1.7	1.3	1.4	1.7	1.8

Thus, if the limit on the number of attempts to interview in the Arab sector is set at three, there will be four attempts, at the most, in the Jewish sector; if a maximum of four attempts is imposed in the Arab sector, the limit in the Jewish sector should be seven. (We also performed a simulation with the in-between limits of four attempts for Arab subjects and five for Jewish subjects. Its results were consistently similar to those in Simulations A and B. Details available from the authors upon request.)

We then compare the results of each simulation with the results of the actual response. Thus, 6,132 of the 7,782 households in the dwelling sample of the 2004 Household Expenditure survey completed the survey. (That is, they completed Part A of the questionnaire, the expenditure log, and Part B of the questionnaire, and were not disqualified in the editing of the data. For further details about the results of the fieldwork, see Central Bureau of Statistics 2006). The number of attempts to interview at the phase of completing Part A of the questionnaire was not limited by the survey management in a uniform way. Instead, the decision about whether to continue the attempts to interview or to desist was made on a case-by-case basis; it ranged from 1 to 22 in the Jewish sector and from 1 to 13 in the Arab sector; the average number of attempts per interviewed subject was 2.67 in the Jewish sector and 1.93 in the Arab sector (Table 4).

Our simulations reflect a situation in which the number of attempts was limited in the full sample and in each sector separately. Consequently, cases which were finally interviewed or identified as not eligible (using more visits than the limit would allow) were now considered as nonrespondents. The stricter the limit, the fewer the cases that were fully processed, and the lower the response rate. Thus, the response rate in Simulation A, where the limit was lower, was 77.2%, whereas in Simulation B, which had a higher limit, it was 85.8%. In the actual survey, which in fact imposed no limit on the number of attempts to interview, the response rate was 88.3%.

Obviously, limiting the number of attempts saves interviewers' time and the organization's budget. The estimates of saved interviewer's time range from 14% in

Table 4. Simulation of Limitation on Number of Interview Attempts, by Sector[a]

	Total	Jews	Arabs
Total households in sample, gross	7,782	6,763	1,019
Actual Results			
Maximum number of attempts		22	13
Noneligible	834	746	88
Fully interviewed	6,132	5,241	891
Response rate, percent, excl. noneligible cases	88.3	87.1	95.7
Total attempts to interview	20,051	18,086	1,965
Average attempts per subject	2.58	2.67	1.93
Simulation A			
Maximum number of attempts		4	3
Noneligible	676	594	82
Fully interviewed	5,488	4,673	815
Response rate, percent, excl. noneligible cases	77.2	75.7	87.0
Total attempts to interview	17,260	15,472	1,788
Average attempts per subject	2.22	2.29	1.75
Simulation B			
Maximum number of attempts		7	4
Noneligible	787	702	85
Fully interviewed	6,002	5,138	864
Response rate, percent, excl. noneligible cases	85.8	84.8	92.5
Total attempts to interview	19,371	17,472	1,899
Average attempts per subject	2.49	2.58	1.86

Source: 2004 Household Expenditure Survey, authors' computations.

[a] Sector of fully interviewed households is defined by religion of head of household. For those not interviewed and for which head of household's religion is not known, it is defined by religion of locality's majority (see Note 2 to Table 3). Since the definition of subject's sector in this table is different from that in Table 1, the proportion of those fully interviewed (of the eligible) is slightly different than that shown in Table 1.

Table 5. Simulation of Limitation on Number of Interview Attempts, by Sector[a]

	Total	Jews	Arabs
Actual results			
1. Fully interviewed	6,132	5,241	891
Simulation A			
2.1 Fully interviewed	5,488	4,673	815
2.2 Total interviewer attempts saved[b]	2,791	2,614	177
2.3 Possible increase in survey sample[c]	1,243	1,143	101
2.4 Expected addition of completed interviews due to increase in sample[d]	870	790	81
2.5 Total expectation of completion of interview [(2.1) + (2.4)]	6,358	5,463	896
2.6 Ratio of expected completion of interview to actual completion of interview [(2.5)/(1)]	1.04	1.04	1.01
Simulation B			
4.1 Completion of interview	6,002	5,138	864
4.2 Total interviewer attempts saved[b]	680	614	66
4.3 Possible increase in survey sample[c]	273	238	35
4.4 Expected addition of completed interviews due to increase in sample[d]	211	181	30
4.5 Total expectation of completion of interview [(4.1) + (4.4)]	6,213	5,319	894
4.6 Ratio of expected completion of interview to actual completion of interview [(4.5)/(1)]	1.01	1.01	1.00

Source: 2004 Household Expenditure Survey, authors' computations.

[a] Sector of fully interviewed households is defined by religion of head of household. For those not interviewed and for which head of household's religion is not known, it is defined by religion of locality's majority (see Note 2 to Table 3).

[b] The difference between total attempts to interview in actuality and total attempts made in a given simulation (Table 4).

[c] The increase in the survey sample due to limiting of attempts to interview, assuming a constant survey budget. The added increment is total attempts saved divided by average attempts per subject in a given simulation (Table 4).

[d] The expectation of completed interviews due to the increase in the sample is the increment added to the survey sample multiplied by the rate of completion of interviews among members of the frame in a given simulation.

Simulation A to 3.5% in Simulation B. If the survey budget is earmarked, one may "reinvest" these savings in increasing the size of the sample. In Table 5, we calculate the increment to the number of completed interviews in the final sample. This addition to the sample is assumed to be interviewed in accordance with the two simulation scenarios. In each simulation the proportion of interviewed subjects will be the same as for the additional cases, as shown in Table 4 (e.g., 75.7 percent in Simulation A in the Jewish sector). With this addition, the expected response rate would surpass the response rate in the actual survey. The increases were 4% in Simulation A and 1% in Simulation B. This result is in accordance with intuition because nonrespondents that require more-than-average attempts to interview would be replaced with average cases, who are less demanding in terms of interview attempts. Hence, one may consider the process of limiting the number of interview attempts as a trade-off between expanding the final sample at the price of a controllable cut of the survey response rate.

Since any effective limit on the number of attempts will lower the response rate, the main question is whether the survey estimates will be biased due to uniqueness of the characteristics of the nonrespondents. To answer this question, we carry the two truncated samples, obtained as a result of limiting the number of interview attempts in the aforementioned simulations, through the standard process of constructing weights, as is customary in the Household Expenditure survey.

The set of weights coefficients is constructed in an iterative procedure known as raking. In this process, the distribution of the weighted sample is adjusted to reflect several external distributions for selected variables – characteristics of households and persons. The adjustment is made for three groups: the population of Jewish and mixed localities, excluding immigrants from 1998 onward; immigrants from 1998 onward; and the population of non-Jewish localities. The household characteristics are household size and age composition (lone elders, young couples, households with children, etc.). The characteristics of individuals are sex and age group, by geographic regions. (For details, see Central Bureau of Statistics 2006).

Through use of the raking process, the estimates of the distribution of these calibrated characteristics should not be affected by the decline in the response rate, provided the process converges. However, there is no assurance that the estimates of other survey variables that are not included in the weight-adjustment process – income, expenditure, schooling, and labor-force characteristics, to name only a few – will remain unbiased as a result of truncating the sample by a limit on the number of interview attempts. As for the outcome of the estimates in terms of precision and bias, two contrasting effects are at work. On the one hand, variance may increase because the samples in the simulations are 11% and 2% smaller than the actual sample of the survey. On the other hand, if the nonrespondents are much different from those who remained in the sample, the estimates may be more precise but biased. Increasing the sample by addition of "average" cases (on account of time saved by foregoing the "hard" cases) is meant to reduce, if not eliminate, the first of these influences. Yet, it should not have any impact on the bias.

Table 6 presents estimates for a range of household characteristics across the full survey sample and the two truncated samples that were obtained in the limited-attempts simulations. First we observe changes in the composition of households as a result of the truncation of the sample. We focus our discussion mainly on Simulation A, which underscores the influence of sample truncation on the quality of the estimates because its limit is the lowest, and the more interview attempts are allowed, the more closely the truncated samples will approximate the actual sample. Thus, in Simulation A the share of large households (four persons or more) increased by 1% at the expense of the share of single-person households – among Jews only, in accordance with the negative correlation of household size to number of attempts needed for interviewing (Table 3).

Before analyzing the weighted estimates of the survey variables, we would mention that the variation coefficient of the weights – in the Jewish sector and generally – declined, i.e., their variance widened, as expected. In contrast, the variation coefficient in the Arab sector did not change in Simulation A but dropped in Simulation B; it may be explained by

Table 6. Estimates of Demographic and Economic Characteristics from the 2004 Household Expenditure Survey – Actual Sample and Two Truncated Samples, by Sector[a] (thousands)

	Actual sample			Simulation A			Simulation B		
	Total	Jews	Arabs	Total	Jews	Arabs	Total	Jews	Arabs
Households in sample	6,132	5,241	891	5,488	4,673	815	6,002	5,138	864
Household composition, non-weighted (pct.)									
1 person		18.5	7.1		17.6	6.6		18.2	7.2
2 persons		26.5	13.6		26.3	13.3		26.4	13.5
3 persons		16.8	11.2		16.9	11.7		16.8	11.1
4 persons		17.6	18.6		18.0	18.8		17.8	18.5
5 + persons		20.6	49.5		21.2	49.6		20.8	49.7
Avg. weight	318.4	315.7	329.2	351.3	350.1	355.6	324.3	320.6	339.2
Weight c.v.	2.066	2.348	1.535	2.007	2.226	1.537	2.040	2.325	1.515
Survey weighted population	6,494.0	5,137.8	1,356.2	6,494.4	5,149.9	1,344.5	6,494.4	5,139.5	1,354.9
Age groups									
0–17	2,146.8	1,544.0	602.7	2,146.7	1,550.6	596.1	2,146.7	1,545.2	601.5
18–24	725.5	556.6	168.9	726.4	556.6	169.8	725.7	555.8	169.9
25–34	1,004.0	796.8	207.2	1,004.0	800.0	204.0	1,004.5	798.2	206.3
35–44	769.6	602.2	167.4	768.5	604.2	164.3	769.6	602.6	167.0
45–54	722.5	621.3	101.3	723.0	621.8	101.2	722.6	621.0	101.6
55–69	698.1	621.3	76.9	698.3	620.7	77.6	698.4	620.9	77.6
70 +	427.4	395.7	31.7	427.6	396.0	31.6	427.0	395.5	31.0
Married	2,721.9	2,224.4	497.6	2,720.7	2,227.3	493.4	2,719.9	2,223.4	496.5
Single	1,067.1	862.7	204.3	1,074.9	871.0	204.0	1,068.8	863.3	205.6
Divorced	211.8	203.9	7.8	207.3	199.5	7.8	212.9	205.0	7.9
Widowed	295.4	258.8	36.6	295.1	258.1	37.0	295.0	258.8	36.1
1989 + immigrants	945.1	875.5	69.6	944.3	878.0	66.2	946.3	875.6	70.7
Secondary education	2,004.0	1,636.2	367.8	2,010.5	1,644.6	365.8	2,007.1	1,638.6	368.5
Post-secondary nonacademic education	705.5	642.0	63.4	694.1	633.4	60.7	704.1	640.8	63.3
Academic education	1,212.8	1,111.8	101.0	1,222.2	1,121.4	100.7	1,211.5	1,110.5	101.0

Table 6. Continued

	Actual sample			Simulation A			Simulation B		
	Total	Jews	Arabs	Total	Jews	Arabs	Total	Jews	Arabs
Labor-force status and sources of income									
Worked (in three months preceding survey)	2,294.1	1,986.4	307.7	2,297.5	1,995.8	301.7	2,295.0	1,987.4	307.6
Employee	2,038.2	1,772.7	265.5	2,052.3	1,787.6	264.7	2,040.7	1,773.4	267.3
Self-employed	255.9	213.7	42.2	245.2	208.2	37.0	254.3	214.0	40.3
Employees, worked <35 hours per week	588.7	522.5	66.2	591.1	525.7	65.5	591.2	525.8	65.4
36–45 hours per week	745.7	657.6	88.1	754.6	666.2	88.4	745.6	656.0	89.6
46 + hours per week	703.9	592.7	110.9	706.6	595.8	110.9	703.8	591.6	112.3
Child-allowance recipients	896.8	702.2	194.6	894.4	700.8	193.6	897.2	702.7	194.6
Disability-benefit recipients	173.8	140.9	32.9	177.5	142.8	34.6	173.8	140.8	33.0
Unemployment-benefit recipients	56.6	46.6	10.0	56.1	46.0	10.2	57.0	46.6	10.4
Income-maintenance recipients	130.4	86.5	43.9	134.3	89.2	45.0	131.2	86.7	44.5
Net income, employee and self-employed, NIS/month	5,580	5,739	4,554	5,581	5,741	4,518	5,584	5,744	4,551
Pension income in Israel, NIS/month	4,556	4,602	3,575	4,465	4,504	3,658	4,560	4,608	3,564
Total expenditure per capita, NIS/month	2,427	2,633	1,648	2,432	2,641	1,633	2,429	2,637	1,640
Total consumption expenditure, NIS/month	10,442	10,627	9,385	10,458	10,653	9,330	10,450	10,642	9,352

Source: 2004 Household Expenditure Survey, authors' computations.

[a] Simulation A: up to 4 attempts to interview Jews, up to 3 attempts to interview Arabs; Simulation B: up to 7 attempts to interview Jews, up to 4 attempts to interview Arabs. Shaded cells represent estimates in which the differences between them and the corresponding estimates in the actual sample are significant at 0.01. Sector of the head of household, "Jews" includes non-Arabs.

unstable raking convergence equilibrium or inadequate design of the Arab subsample, as discussed below.

General observation of the estimates of the traits of the survey population shows that the estimates for the Arab sector are less robust in both simulations, although the differences relative to the actual sample are rarely statistically significant. Since the sample in Simulation B is only 3% smaller than the actual sample, one may infer that lack of robustness in the estimates (a typical feature of both simulations) has nothing to do with the limit that was imposed on the number of interview attempts; instead, it evidently originates in the small sample of Arabs.

In the Jewish sector, the simulations rarely biased the estimates by more than 1% relative to the actual sample. Statistically significant biases were observed in the estimates of only two variables: number of self-employed (in both simulations) and pension receipts (in Simulation A). These populations are known for their less-than-complete coverage in the Household Expenditure Survey. For instance, according to the income tax records, there were 307,000 active self-employed persons in Israel in 2004, as against 256,000 according to the estimate in the actual sample.

When the cases in the survey population are counted and compared with administrative data from the income tax authorities and the National Insurance Institute, underreporting problems turn up every year, both in the population of recipients and in the benefit receipts and income of the self-employed. Since administrative records are able to provide these data, the best way to improve the estimates, we believe, is by matching to these sources and imputation of administrative data. In any case, it should be borne in mind that the estimations in our simulations were based on samples that were smaller than the actual one (Table 6).

Finally, we computed the mean squared error (MSE), presented in Table 7 as a ratio to the simulation estimates. For each variable, the MSE was computed around the respective estimate in the actual sample, i.e., we measure only the bias induced by the sample truncation. As a rule, the results are consistent across the simulations: MSE is higher in Simulation A, with the lower limit on the number of interview attempts. For the Jewish sector and in total population, a few variables have a relative MSE that passes the 5% level. A group of three variables constitutes a noticeable exception: recipients of the disability, income-maintenance and unemployment benefits, whose relative MSE is well above 5%. According to the data of the National Insurance Institute (which pays these benefits), the average number of recipients in 2004 stood at 162.4, 145.1, and 58.7 thousand, respectively, in the survey population of 4.3 million aged 18 and above (compared to the estimates of 173.8, 130.4, and 56.6 thousand, respectively, in the actual HES sample). The relative MSE goes down quite slowly from Simulation A to Simulation B. That is, like in the case of the self-employed estimates, we would urge treating the bias by using the available administrative data.

For the Arab sector, numerous lines in Table 7 exhibit two-digit figures, reaching 25% in the estimates of the divorced and the unemployment-benefit recipients. Again, it seems to be a problem related to estimation of the small population subgroups in a small sample, rather than to truncation of 8.5% to 3% of the sample of Arabs, since the two-digit figures do not disappear even in the mildly truncated sample in Simulation B.

Table 7. Relative Root Mean Squared Error of Demographic and Economic Characteristics from the 2004 Household Expenditure Survey – Two Truncated Samples, by Sector[a] (percent)

	Simulation A			Simulation B		
	Total	Jews	Arabs	Total	Jews	Arabs
Age groups						
0–17	1.9	2.2	4.6	1.8	1.9	4.4
18–24	3.0	3.3	6.8	2.8	3.1	6.5
25–34	2.6	2.9	6.3	2.4	2.7	6.1
35–44	2.6	2.9	6.1	2.4	2.7	5.7
45–54	2.8	3.0	7.9	2.7	2.8	7.7
55–69	2.9	3.0	9.5	2.7	2.8	9.1
70 +	3.2	3.3	14.0	3.1	3.1	14.3
Married	1.4	1.6	3.7	1.3	1.4	3.6
Single	2.8	3.0	7.4	2.6	2.7	7.2
Divorced	5.2	5.2	25.6	4.5	4.6	24.3
Widowed	3.7	3.9	13.4	3.5	3.6	13.2
1989 + immigrants	3.2	3.4	13.3	3.0	3.2	12.7
Secondary education	1.9	2.1	4.9	1.8	1.9	4.8
Post-secondary nonacademic education	3.2	3.1	11.4	2.6	2.7	10.6
Academic education	2.3	2.4	8.5	2.0	2.1	8.3
Labor-force characteristics and sources of income						
Worked (in three months preceding survey)	1.5	1.7	5.0	1.4	1.5	4.5
Employee	1.8	1.9	5.0	1.5	1.6	4.9
Self-employed	5.8	5.0	16.0	4.1	4.3	12.5
Employees, worked < 35 hours per week	2.8	3.0	9.1	2.7	2.8	9.1
36–45 hours per week	2.8	3.0	8.2	2.4	2.5	8.1
46 + hours per week	2.6	2.8	7.4	2.4	2.5	7.3
Child-allowance recipients	1.7	1.9	4.2	1.6	1.8	4.1
Disability-benefit recipients	6.1	6.5	14.3	5.4	6.0	11.8
Unemployment-benefit recipients	9.5	10.4	23.7	9.1	9.8	24.0
Income-maintenance recipients	7.0	8.2	12.0	6.0	7.0	11.1
Net income, employee and self-employed, NIS/month	1.0	1.0	2.6	0.9	1.0	2.4
Pension income in Israel, NIS/month	3.2	3.3	17.2	2.7	2.7	15.8
Total expenditure per capita, NIS/month	1.1	1.2	3.3	1.0	1.1	2.9

Source: 2004 Household Expenditure Survey, authors' computations.

[a] Simulation A: up to 4 attempts to interview Jews, up to 3 attempts to interview Arabs; Simulation B: up to 7 attempts to interview Jews, up to 4 attempts to interview Arabs. Sector of the head of household, "Jews" includes non-Arabs, Relative Root MSE $= \sqrt{\left(s_{\hat\theta}^2 + (\hat\theta - \theta)^2\right)/\hat\theta}$, where θ is the estimate in the actual sample, $\hat\theta$ is the estimate in the truncated sample.

Given the results of the simulation, we conclude that limiting the number of interview attempts in the Israeli Household Expenditure survey to three in the Arab sector and four in the Jewish sector would not bring about a considerable nonresponse bias in the key survey variables. At the same time, the robustness of the estimates could be improved markedly by the usage of available administrative data and increase of the sampling fraction in the Arab sector.

5. Conclusion and Discussion

Face-to-face interviewing is the most expensive method that a statistical organization can use to survey. Therefore, organizations around the world are seeking more economical alternatives. Telephone interviewing or administrative information sources are possible solutions that may help (e.g., Lebrasseur and Dion 2005), but they are not always readily available. In most countries the major surveys, such as Labor Force Surveys and Households Expenditure or Income Surveys, continue to be performed using face-to-face interviewing. Therefore, the streamlining of fieldwork, including setting rules as to number of attempts to interview remains a significant issue.

Some may claim that the remuneration of interviewers by output (number of subjects interviewed) absolves the organization of the need to limit the number of interview attempts, because it shifts responsibility for deciding how many attempts are worthwhile onto the interviewers. In order to work effectively an interviewer needs to strike a balance between the circumstances of each subject in their workload, and the probability of success in view of a predetermined level of response. This solution, however, is inapplicable when the interviewers are civil servants, who are not customarily paid on a piecework basis. Moreover, paying by output may encourage interviewers to concentrate their efforts on interviewing the easier subjects, which in turn might contribute to nonresponse bias due to under-representation of subjects who are difficult to interview.

To resolve this matter, some national organizations impose limits on the number of attempts to interview. The decision as to the limit is usually dependent mostly on budget constraints. In this article, we present a cost-benefit model in which the optimal maximum number of attempts is determined at the intersection of the marginal cost of attempts and the expected utility of the information. Beyond this number of attempts, the interviewing effort is not economically justified.

Using multivariate regression analysis, we show that several variables influence the number of attempts needed to interview. These include variables that are known before the interviewing process begins, such as locality size, subject's sector (Jew or Arab), and interviewer characteristics, and variables that become known from the questionnaire, including household size, age of head of household, and employment status of the household members. Altogether these findings show that the probability of interviewing varies among subjects, suggesting the possibility of setting different limits on the number of interview attempts in accordance with these variables.

To demonstrate the point, we applied the model in two simulations on the basis of Israel's 2004 Household Expenditure Survey. In the simulations, the number of interview attempts was limited and the main survey estimates were recalculated on the basis of the truncated survey sample. We compared these estimates with corresponding estimates from the actual sample, in which no effective limit was set on the number of interview attempts. Obviously, imposing a limit on the number of interview attempts caused a decline in the response rate in the truncated samples. This decline, however, did not introduce a statistically significant bias in the main survey variables – demographic characteristics, labor-force characteristics, personal income, and household expenditure by main consumption groups. Even with the lowest simulated limit on interview attempts, the survey estimates displayed a great deal of robustness, except for two variables – the number of self-employed persons and pension recipients.

Since these two variables are known to be underreported in the actual sample as well, in which the number of interview attempts is not limited, we conclude that sample truncation did not cause the problem but merely confirmed its existence. The simulation estimates for the Arab sector are generally less robust and do not correspond to the limitation of attempts, thereby pointing to the need to increase the survey sample in this sector.

The upshot of this discussion is that the number of interview attempts in the Household Expenditure Survey may be held to a low limit of three in the Arab sector and four in the Jewish sector. This would lower the response rate from above 88% to 77% (76% of Jews, 87% of Arabs) and would save about 14% of the interviewers' time, without biasing the survey estimates, provided that the sample fraction in the Arab sector is increased.

While considerable savings may be achieved by making fewer interview attempts, the effect of such an action on the management of fieldwork should be pondered thoroughly, especially if the limitation of interview attempts is to be differentiated on the basis of the characteristics of the subjects and/or the interviewers. So far, the ICBS fieldwork staff has been instructed that every effort must be made to maximize the response rate. This message is embodied in the pay system (premium for surpassing the predetermined response-rate goal of 87%) and is stressed repeatedly by supervisors when controlling and evaluating the interviewers' work; it resonates as the leitmotif in training sessions. Limiting the number of interview attempts could be very confusing and might impair the motivation of the staff involved. Therefore, before implementing the model it would be necessary to introduce considerable changes in several aspects of the fieldwork, such as interviewers' control and evaluation methods.

The practical introduction of the model depends, *inter alia*, on the efforts that the interviewer has to invest in each interview attempt. Further, factors such as geographical dispersion of the interviewer's workload make some visits more demanding than others. It is important to consider whether an across-the-board limit on the number of attempts will be useful even in cases where the attempt does not entail much time or effort, such as when the subject is on the interviewer's route in any case.

The findings pointing to interviewer's characteristics which tend to reduce the number of attempts may lead to conclusions about revising the process used to screen and hire interviewers. In practice, it is quite difficult to recruit interviewers; it is not clear whether an organization can in fact hire highly educated people for this job. Furthermore, there are requirements that impede the identification of appropriate interviewers, including fluency in languages other than Hebrew (Arabic, Russian). Under these circumstances, improving the training process seems to be the best available way to enhance the interviewers' abilities and motivation.

Additional difficulties of an administrative nature arise if a differential number of attempts to interview is set in accordance with sample composition, e.g., the Arab and Jewish sectors, as we did in the simulation. In this case an organization will have to take at least one of the following two actions: to train interviewers to make decisions about investing resources in accordance with sampling unit characteristics that are disclosed to them in advance, or to assure appropriate instructions from the coordinators for each sampling unit. Each of these actions entails a new array of control and supervision.

It may be instructive to consider the decision to limit the number of interview attempts in the framework of multiple phase *responsive design* (Groves and Heeringa 2006), which suggests monitoring and adjusting fieldwork procedures during the fieldwork process. The maximum number of attempts should be determined in the planning stage, based on the suggested cost-benefit model. The first phase of the fieldwork process would be complete when interviewers either succeed in interviewing or reach the limit of interview attempts without completing the interview. At this point the decision has to be made regarding the mode of fieldwork operation in the second phase. The options for the second phase include: (i) adding units to the original sample in order to substitute nonrespondents of the first phase; (ii) allowing interviewers to make more attempts to interview – beyond the predetermined limit; and (iii) finalizing the nonresponse result. These options should be weighted based on information (paradata) collected before and during the fieldwork, and the marginal cost of a visit to a specific subject (group of subjects). However, it may matter, from the methodological point of view, if in a given survey population noncontacts are different from refusals and omitting them by the same proportion (under a uniform limit on the number of attempts) would result in a disproportional bias. Then, it could be desirable to set different interview attempts limits for the two groups. Evidently, this is a particular case to our model, just like setting the different limits for Jews and Arabs, as considered in the simulations, or for urban and rural populations, or for any other partition of the sample on groups that deserve different treatment. At any rate, setting different limits and coordinating interviewers workloads in accordance with them have to be embedded effectively in the flow of the fieldwork within the survey timetable.

Despite the complications of introducing new procedures in an ongoing survey, the model's implementation can save the organization scarce resources, that can be reinvested in other measures that are needed for the surveying work, e.g., increasing the sample size or improving performance by means of instructional and training activities.

Before concluding, a few reservations should be mentioned. First, our study compared two truncated samples to the actual sample in order to evaluate a possible nonresponse bias, caused by limiting the number of attempts to interview. There still might be bias in the survey's actual sample, in some important variables. We did not verify this, having examined only a limited set of (key) variables reported by all subjects in the sample. Using external information, e.g., from a benchmark survey on the same subjects, like an updated census, or administrative sources, could have shed more light on the biases in the relevant survey variables.

Second, it has been argued that truncating the sample in order to simulate lower response rates resulting from fewer visits (Curtin et al. 2000; Heerwegh et al. 2007) is different than designing a survey which limits interviewers' efforts. Given a predetermined limit, interviewers might have changed the way they plan and organize their trips, and their timing, to make them more effective – leading to different results. Nonetheless we believe our results give a fairly good approximation of the results obtained with less visits. As mentioned elsewhere, most visits are scheduled in the evenings, which we found to be most effective. After an initial evening visit, interviewers are required to schedule their visits for different times during the day and for different days, and when possible, base their schedule on information given by the household or neighbors. This procedure is well monitored in order to insure maximum effectiveness of the fieldwork.

Even if interviewers were instructed to limit their visits, we doubt they could change their schedule substantially to make it more effective.

Third, the response rate in Israel's Household Expenditure Survey is very high (around 88%) relatively to similar surveys elsewhere. Applying the suggested model could have yielded different results as to a possible bias caused by limiting the efforts for surveys with lower response rates.

Finally, several idiosyncratic features of the HES and its fieldwork must be taken into account when thinking of generalizing our model and results to other surveys. These include the fact that the survey is mandatory, as stated in a preliminary letter sent to all sampled households. The collection method – face-to-face interviews – is enormously expensive, causing a constant need for controlling and monitoring the number of visits in order to ease their heavy burden on the survey's budget. The HES interviewers are generally well-trained and experienced, so they might have developed efficient work procedures to reduce the visits needed before being able to interview. The survey enjoys a high level of cooperation among the respondents because it provides data for computation of the poverty line and the consumer price index. In other surveys with different features, the bottom line of the cost-benefit analysis and a resulting limit on interview attempts (visits or telephone calls) could be very different.

Therefore, more surveys should be examined in order to check the findings, so that the policy on enhancing the efficiency of fieldwork, and the requisite preparations for deploying such a policy, would encompass the widest possible range of different surveys.

6. References

Abraham, K.G., Maitland, A., and Bianchi, S.M. (2006). Nonresponse in the American Time Use Survey: Who Is Missing from the Data and How Much Does It Matter? Public Opinion Quarterly, 70, 676–703.

Central Bureau of Statistics (2006). Household Expenditure Survey 2004, General Summary. Special Publication 1261. Israel: Jerusalem.

Curtin, R., Presser, S., and Singer, E. (2000). The Effects of Response Rate Changes on the Index of Consumer Sentiment. Public Opinion Quarterly, 64, 413–428.

Durrant, G.B. and Steele, F. (2007). Multilevel Modelling of Refusal and Noncontact Nonresponse in Household Surveys: Evidence from Six UK Government Surveys. University of Southampton, Southampton Statistical Sciences Research Institute, S3RI Methodology Working papers, M07/11.

Feskens, R., Hox, J., Lensvelt-Mulders, G., and Schmeets, H. (2007). Nonresponse Among Ethnic Minorities: A Multivariate Analysis. Journal of Official Statistics, 23, 387–408.

Greene, W.H. (2000). Econometric Analysis. New Jersey: Prentice-Hall, Inc.

Groves, R.M. (2004). Survey Errors and Survey Costs. New York: John Wiley and Sons.

Groves, R.M. (2006). Nonresponse Rates and Nonresponse Bias in Household Surveys. Public Opinion Quarterly, 70, 646–675.

Groves, R.M. and Couper, M.P. (1998). Nonresponse in Household Interview Surveys. New York: John Wiley and Sons.

Groves, R.M., Couper, M.P., Presser, S., Singer, E., Tourangeau, R., Acosta, G.P., and Nelson, L. (2006). Experiments in Producing Nonresponse Bias. Public Opinion Quarterly, 70, 720–736.

Groves, R.M. and Heeringa, S.G. (2006). Responsive Design for Household Surveys: Tools for Actively Controlling Survey Errors and Costs. Journal of the Royal Statistical Society, Series A, 169, 439–457.

Groves, R.M., Presser, S., and Dipko, S. (2004). The Role of Topic Interest in Survey Participation Decisions. Public Opinion Quarterly, 68, 2–31.

Groves, R.M., Singer, E., and Corning, A. (2000). Leverage-saliency Theory of Survey Participation. Description and An Illustration. Public Opinion Quarterly, 68, 2–31.

Heerwegh, D., Abts, K., and Loosveldt, G. (2007). Minimizing Survey Refusal and Noncontact Rates: Do Our Efforts Pay Off? Survey Research Methods, 1, 3–10.

Hox, J. and de Leeuw, E. (2002). The Influence of Interviewers' Attitude and Behavior on Household Survey Nonresponse: An International Comparison. Survey Nonresponse, R.M. Groves, D.A. Dillman, J.L. Eltinge, and R.J.A. Little (eds). New York: John Wiley and Sons, 103–120.

Kiefer, N.M. (1988). Economic Duration Data and Hazard Functions. Journal of Economic Literature, 26, 646–679.

Lebrasseur, D. and Dion, S.M. (2005). The Telephone First Contact Approach in the Labour Force Survey. Proceedings of Statistics Canada's Symposium "Methodological Challenges for Future Information Needs".

Lynn, P. and Clarke, P. (2002). Separating Refusal Bias and Non-contact Bias: Evidence from UK National Surveys. The Statistician, 51, 319–333.

Lynn, P., Clarke, P., Martin, J., and Sturgis, P. (2002). The Effects of Extended Interviewer Efforts on Nonresponse Bias. Survey Nonresponse, R.M. Groves, D.A. Dillman, J.L. Eltinge, and R.J.A. Little (eds). New York: John Wiley and Sons, 135–148.

Nicoletti, C. and Peracchi, F. (2005). Survey Response and Survey Characteristics: Microlevel Evidence from the European Community Household Panel. Journal of the Royal Statistical Society, Series A, 168, 763–781.

Olson, K. (2006). Survey Participation, Survey Nonresponse Bias, Measurement Error Bias, and Total Bias. Public Opinion Quarterly, 70, 737–758.

O'Muircheartaigh, C. and Campanelli, P. (1999). A Multilevel Exploration of the Role of Interviewers in Survey Non-Response. Journal of the Royal Statistical Society, Series A, 162, 437–446.

Philippens, M., Billiet, J., Loosveldt, G., Stoop, I., and Koch, A. (2004). Non-Response and Fieldwork Efforts in the ESS: Results from the Analysis of Call Record Data. Work Package 7, Data-Bases Quality Assessment in the ESS. Available at http://naticent02.uuhost.uk.uu.net/methodology/nonresponse_fieldwork_efforts.pdf

Purdon, S., Campanelli, P., and Sturgis, P. (1999). Interviewers'Calling Strategies on Face-to-Face Interview Surveys. Journal of Official Statistics, 15, 199–216.

Schechtman, E., Yitzhaki, S., and Artsev, Y. (2008). Who Does Not Respond in the Household Expenditure Survey: An Exercise in Extended Gini Regressions. Journal of Business and Economic Statistics, 26, 329–344.

Shinar, S., Oren, O., and Levin-Epstein, N. (2005). Response to Surveys in Israel. Cohen Institute News, no. 3, Tel Aviv University, The Gershon Gordon Faculty of the Social Sciences (Hebrew).

Slemrod, J. and Yitzhaki, S. (2002). Tax Avoidance, Evasion and Administration. Handbook of Public Economics, A.J. Auerbach and M. Feldstein (eds). (First Edition). Vol. 3. Amsterdam: Elsevier, 1423–1470.

Stoop, I. (2004). Surveying Nonrespondents. Field Methods, 16, 23–54.

Sturgis, P., Smith, P., and Hughes, G. (2006). A Study of Suitable Methods for Raising Response Rates in School Surveys. UK Department for Education and Skills, Research Report No. 721, BMRB International Limited.

Teitler, J., Reichman, N., and Sprachman, S. (2003). Costs and Benefits of Improving Response Rates for a Hard-to-Reach Population. Public Opinion Quarterly, 67, 126–138.

Received May 2007
Revised October 2009