



**Economic and Social
Council**

Distr.
GENERAL

ECE/CES/2007/35
11 May 2007

ENGLISH ONLY

ECONOMIC COMMISSION FOR EUROPE

STATISTICAL COMMISSION

CONFERENCE OF EUROPEAN STATISTICIANS

Fifty-fifth plenary session
Geneva, 11-13 June 2007
Item 5 of the provisional agenda

SEMINAR ON INCREASING THE EFFICIENCY AND PRODUCTIVITY
OF STATISTICAL OFFICES
SESSION II

Get it or drop it? A model for cost-benefit analysis of the enumeration process in household surveys

Submitted by Central Bureau of Statistics, Israel

ABSTRACT

This paper develops a cost-benefit model with which an optimal allocation of resources for reaching for a respondent may be determined. The model accounts for heterogeneous response propensity among various population groups, the costs of fieldwork, and the utility of the information gathered in the survey. Putting a cap on the number of attempts to interview lowers the survey's response rate because it forgoes the coverage of hard-to-reach populations and conversion of refusals, possibly creating a non-response bias—a factor not directly reflected in a cost-benefit reckoning. We demonstrate the use of the model by simulating different caps 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 87% to 76%—we find that limiting the number of attempts to interview causes no significant bias in estimates of the main survey variables.

Key words: fieldwork efficiency, non-response bias, cap on attempts to interview.

INTRODUCTION

1. It is widely believed that the more attempts to interview are made, the higher the response rate and the better the survey estimates will be. However, the budget constraints that apply to any survey make it necessary at some point to forego hard-to-reach populations. Often, the surveyors' decisions about where and when to desist from further attempts to interview unresponsive sample units are arbitrary and ad hoc relative to the circumstances of the case at hand. Thus, the present study asks whether it is possible to analytically devise an optimum number of attempts to interview as an across-the-board rule for surveys of a given type while minimizing the possible bias that this would inflict on the survey estimates.

2. Numerous studies have proven that hard-to-reach populations differ from those that are more easily interviewed. Among hard-to-reach populations, a distinction is made between sample units who are not interviewed due to difficulty in establishing contact (non-contacts) and those who refuse to participate in the survey (refusals).¹ Philippens et al. (2004), Teitler et al. (2003), and Lynn et al. (2002) show that interviewers have to make extra exertions to interview these two population groups, i.e., the higher the desired response rate, the higher the marginal cost of enumeration. The risks of non-response bias in the survey estimates are usually cited to justify the need to raise the response rates. However, as Sturdis et al. (2006) note, raising the response rate does not assure survey quality:

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 that fails to meet the criteria. More emphasis should be placed, where possible, on bias assessments that compare responders to non-responders on key survey variables (p. 2).

3. Furthermore, not every decline in response rate takes a severe toll on survey quality. For example, the recent study of Abraham et al. (2006) on the American Time Use Survey, in which the response rate was under 60%, shows that survey estimates calculated with the base weights unadjusted for non-response, are broadly similar to those obtained by means of two sets of weights adjusted for non-response. Groves (2006), considering the linkage between non-response rates and non-response bias, concludes that recent empirical findings seemingly defy conventional wisdom of strong causality between the two factors.

4. While the practical utility of a higher response rate in terms of survey quality may be small if not negligible, 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 complexity of the survey design. A survey of the socio-demographic and behavioral characteristics of fathers among mostly unwed parents (Teitler et al., 2003) found that reaching for subjects whose interviewing required an especially large resource investment did not prevent non-response bias. Furthermore, the characteristics of especially "expensive" sample units resembled those on whom medium resources were invested and not those who were not interviewed. Even though these results cannot be

¹ For a survey of the recent literature, see Groves and Couper (1998), Groves et al. (2002), Nicoletti and Peracchi (2005).

extrapolated to other surveys, the authors of the study urge survey organizations to carefully examine the amount of resources invested in raising the response rate. 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 (2004) barely addressed the *dual problem* that bothers almost every statistical organization in the world: what are the ways and means of marginally economizing on a survey with a given design, without magnifying the survey errors?

5. Although national statistical organizations are publicly funded and therefore subject to stringent budget constraints, none—to our knowledge—uses an economic model that would provide a rational rule for determining resource allocation and how much effort should be invested in fieldwork for different surveys among different population groups using different collection methods.

6. This study develops a cost-benefit model from which one may derive an optimum number of attempts to interview in household surveys performed by personal interviews, such that attempts beyond this number are not worth making. The importance of the model is its applicability in a wide variety of the most expensive surveys—which include personal or telephone interviews. The model may also be used to establish within one survey differential enumeration rules for population groups that are expected to exhibit different response rates.

7. To demonstrate the use of the model, we carry out three simulations with different caps on the number of attempts to interview in the Household Expenditure Survey (hereinafter: HES) of Israel's Central Bureau of Statistics. We also examine the effect of type of locality, population group, and interviewer's effect on the number of attempts to interview by estimating the hazard rates of reaching a potential respondent. Within these factors, we distinguish between those that are known before the interview, which may be useful in managing the fieldwork and planning the enumeration efforts, and those that become known only from the interview, which are more important for methodological research.

8. Israel's HES is performed in three phases. In Phase 1, the interviewer sets out to locate the sample unit's home and present the survey to the household. In this phase, she attempts to persuade the members of household to fill in Part A of the questionnaire, which solicits demographic details about all members of household. In this visit, the interviewer also asks the members of the household to keep an expenditure diary for a two-week period. The number of attempts to interview in this phase includes attempts to locate the household, to persuade its members to take part in the survey, and to interview the household members in order to complete Part A of the questionnaire. In Phase 2, which lasts about two weeks, the members of the household are visited regularly (at least four times, according to the survey guidelines, unless asked by the household not to come) as they make entries in the expenditure diary. In each visit, the interviewer examines the record of expenditures in the diary and encourages them to continue recording all their expenditures. In Phase 3, the interviewer makes a final visit and completes the record of expenditures by eliciting a detailed account of large or irregular outlays for a variety of goods and services that were not covered in the diary; she also completes Part B of the questionnaire, which covers the household's durable goods, income, and work specifics. As a rule, the same interviewer performs all three phases of the HES; this practice is believed to be crucial for establishing good working

relations with the members of the household. The result is a response rate of more than 85% 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. In this study, we focus on the number of attempts to interview in Phase 1 of the survey, because this is the phase that resembles a typical face-to-face survey.

9. The rest of this study is organized in the following way: Section 2 presents the cost-benefit model of the enumeration process, from which we derive a rule for determining an optimum number of attempts to interview. In Section 3, we estimate the effect of the characteristics of interviewer and household on the number of attempts to interview in the HES Phase 1. Section 4 presents a simulation of various caps on the number of attempts to interview and compares the survey estimates from the actual sample with those that were received from the truncated samples with lower response rates. This comparison serves to gauge any non-response bias that the cap on the number of attempts to interview may have created. The final section discusses issues related to the application of the proposed model in household surveys' fieldwork.

I. THE MODEL

10. The schematic model that follows describes the cost and benefit of attempts to interview respondents in a household survey. It may pertain to a survey in which a face-to-face interview should be performed with all or one of the household members at the sampled address (in a survey framed by a dwellings register), or with a specific person (in a survey framed by a population register). In this general setup, the interviewer has to visit the home of the sampled household (person) at least once; if she fails to interview in the first attempt, further attempts are made again and again by additional visits of the interviewer.

I.1 The Cost-Benefit Model

11. The organization's variable of choice is the number of the attempts to interview until the interview is accomplished. Let $n_{ij} = 1, \dots, N_{ij}$ denote the number of attempts to interview until interviewer j enumerates sample unit i . We assume that the "quality" of each attempt by a given interviewer is constant and that the interviewer always makes her best effort to accomplish the interview.

12. The final result of the enumeration effort is either a success (completed interview), $R_{ij} = 1$, or a failure (non-interviewed subject), $R_{ij} = 0$. For simplicity's sake, let us establish that only the interviewer's efforts, as measured by the number of attempts, affect the enumeration result. The effects of observable and non-observable characteristics of interviewer and subject on the likelihood of enumeration are discussed within the context of an econometric model specification.

13. Without a loss of generality, let us assume that the marginal cost of each attempt to interview is constant and define it as c .² The marginal cost includes all variable costs (travel

² The marginal cost of an attempt to interview may not be fixed, depending on how the survey's fieldwork is organized. When an interviewer in the HES is given a list of addresses to attempt to interview (the "workload"), in most cases these are planned to be geographically close to each other. As more and more attempts to interview are made, the number of non-interviewed subjects in the workload declines and they become more dispersed

and searching time, transport, communication, and, when the respondent is interviewed, interview time) and the corresponding share of all relevant fixed costs.³

14. The social utility of enumerating the sample unit, B_i , expresses the social value of the information that the respondent is asked to provide. Since participation in surveys is a sort of a “lump-sum tax” that is imposed on randomly selected sample units, the social utility is equal for every respondent ($B_i = B$ for all i) irrespective of the number of interview attempts needed.⁴ Since the utility is realized only if the survey questionnaire is completed, the expected utility is defined as:

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

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

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

The marginal probability of enumeration, $f(n) \equiv F'(n)$, is a decreasing function of the number of attempts to interview, i.e., $f'(n) < 0$. Therefore, the expected marginal utility declines with the number of attempts to interview: $dE'(B)/dn < 0$.

Thus, by expressing the utility in monetary terms and juxtaposing it to the cost of enumeration, we may determine an optimum number of attempts to interview. In Figure 1, this number is noted at point n^* : every attempt to interview up to this point will deliver, in expectation, positive utility (net of cost).

15. Figure 1 relates to a situation in which the probability of enumeration and, in turn, the expected enumeration utility are uniform for the entire survey population. In practice, this does not happen due to heterogeneity in response propensity among different groups, e.g., Jews and Arabs, urban and rural localities, and, possibly, neighborhoods within one town that are differentiated by socioeconomic level, religion, etc. Accordingly, Figure 2 shows, as examples, two marginal expected utility curves. Type B_1 subjects are “easy” to interview. A large majority of the sample is interviewed in the first attempt, after which the enumeration rate falls steeply. In contrast, Type B_2 subjects require, in expectation, more attempts to interview, as reflected in a thicker and longer right-hand tail of $f(n)$. Accordingly, these traits should be taken into account in determining the optimum number of attempts to interview,

geographically. Thus, the cost of reaching each non-interviewed subject rises. At a certain point, the interviewer is given another workload in a different geographic area. From then on, she has to travel even a greater distance to interview subjects from the previous workload. However, the need to reach out for subjects in the previous workload area in phases 2 and 3 of the HES reduces somewhat the number of extended attempts to interview to the residual workload. Furthermore, it should be borne in mind that the cost of the first attempt to interview is higher than that of subsequent attempts because only on the first occasion does the interviewer have to locate the subject's address.

³ This *administrative* cost to the government (represented in this case by the Central Bureau of Statistics) is smaller than the *social* cost of the enumeration. The latter also includes the subject's response burden, which is a positive function of interview duration, frequency of contacts with subject in a given survey, and the psychic cost to the subject. From these components one should offset the utility to the subject of the information that she gains as a result of the survey. For example, participation in a household-expenditure survey gives one an accurate picture of one's expenses and income.

⁴ The social utility of the respondent's enumeration may depend on the number of attempts to interview at the phase at which questionnaires are completed; we discuss this issue below. If, for example, a certain group of respondents is strongly inclined to attrition at the phase of filling in the expenditure diary—a phase that requires more frequent interviewer visits—the information to be gathered from this group may have greater social value than that of other groups that cooperate more easily.

$n_1^* < n_2^*$. Otherwise, setting a single standard optimum number of attempts to interview for both groups would result in excessive attempts to interview the Type B_1 respondents and failure to fulfill the enumeration potential of the Type B_2 respondents.

16. To use the same model for different types of household surveys, it is necessary to adjust the “scale of measurement” due to the substantial difference between proxy and personal surveys in terms of the enumeration process. In a proxy survey, the enumeration process succeeds when the interviewer gets her “foot in the door” and interviews any adult member of the household, who provides information about all members of the household. The situation is different in personal surveys: the interviewer must interview a specific person who has been targeted for investigation. Therefore, personal surveys have usually a higher rate of non-response due to outcomes such as “subject not home,” “not available for interview,” “communication problems,” etc., than proxy surveys.

17. Consequently, any number of attempts to interview will deliver a lower probability of enumeration in a personal survey (represented in Figure 3 by Curve $E'(B_2)$) than in a proxy survey (Curve $E'(B_1)$). If we were to apply the model without keeping this fact in mind, i.e. if we used the same “scale of measurement,” we would always obtain a smaller optimum number of attempts to interview in a personal survey than in a proxy survey among the same population: $n_1^* < n_2^*$. This will cause, in turn, a much lower response rate in a personal survey, for two reasons: (i) a lower marginal enumeration probability for each number of attempts to interview and (ii) a smaller optimum number of attempts. Therefore, the optimum number of attempts to interview in personal surveys should be determined in accordance with a higher “scale of measurement” than that used in proxy surveys, in order to take into account the difference between the enumeration processes. The technique for applying the model, discussed later in this section, responds to this imperative.

I.2 Application of the Model

18. To apply the model, three kinds of data are needed:
- (1) the marginal cost of an attempt to interview (c);
 - (2) an estimate of the marginal enumeration probability function ($f(n)$);
 - (3) a monetary estimate of the social utility of subject enumeration (B).

The **marginal cost of an attempt 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 parameter.

The **marginal enumeration probability function** is estimated, for each survey and in accordance with the types of subjects in the survey, by duration data analysis models. If N is the number of attempts needed to interview a given subject (or, if the population includes different types of subjects, a subject of a certain type), then the survival function is defined, in terms of number interview attempts, as the probability of non-interview before attempt n :

$$S(n) \equiv \Pr(N > n) = 1 - F(n) \quad (3)$$

19. In the next section, we estimate the survival function by means of a non-parametric analysis (the Kaplan-Meier estimator). We also estimate a parametric model of the accelerated failure time type:

$$\Pr(N > n | X) = \Pr(N_0 > \exp(-X' \beta)n) \quad (4)$$

where X is the array of factors that explain enumeration-process duration and N_0 is the baseline event time, calculated in view of a certain distribution for $X=0$. The relevant distribution is chosen by examining the correspondence between the data and several theoretical distributions; we chose the exponential distribution specification. The array of X can be partitioned into variables that are known before and after enumeration. The first group includes the characteristics of the interviewer and of the respondent's residential environment.⁵ The second group includes all values reported by the respondent in the interview; obviously, they are known only for respondents who complete the interview. Since the average interviewer's workload in the Household Expenditure Survey is about 100 households, the model may be estimated with interviewers' fixed effects that would presumably reflect their "trade secrets," motivation, and other unobserved traits.

The **monetary estimate of the social utility of subject enumeration** is unfortunately immeasurable. As a proxy, one may suggest the time that the subjects need to complete the survey questionnaire—interview length, T —valued at the average value of a leisure hour, C_T : $\hat{B} = T \cdot C_T$.⁶ However, there is a conceptual difficulty in estimating C_T . In the economic literature, there is no consensus about the valuation of the leisure time that the respondent forfeits when participating in a survey. Furthermore, even if a valuation of leisure time could be obtained in some way, it would be an underestimate of the *social utility* of the information elicited by the survey because this information is a public good that has a positive externality effect. Therefore, a reasonable monetary estimate of B cannot be obtained, making it impossible to arrive at an *absolute* determination of the optimum number of attempts to interview in any survey.

20. However, as long as it is possible to approximate the social utility of enumeration in survey k in the form of $B_k = T_k C_T \mu$, where $\mu > 1$ is the multiplier that corrects for the positive-externalities effect of the information,⁷ one may construct a social-utility ratio between any pair of surveys: $B_1/B_2 = T_1/T_2$. The transformation that expresses social utilities relative to response time rests on the assumption that the production of all surveys by statistical organizations is efficient, at margin, in terms of the response burden. In other words, this assumption says that the length of the questionnaire and the burden on the respondent accurately reflect society's need for the information gathered in the survey at issue. The marginal-efficiency requirement pertains only to the contents and length of the questionnaires⁸; this assumption definitely seems reasonable in view of the efforts usually

⁵ In a dwelling survey (such as the Household Expenditure Survey), a subject's neighborhood of residence usually attests to her nationality, for in Israel the vast majority of Jews and Arabs live in segregated neighborhoods. In a personal sample, in addition to the subject's address, h/her basic demographic attributes—sex, age, nationality and religion, composition of nuclear family—are known from the Population Register, which constitutes a framework for the sampling. Border-control data indicate whether the subject was in the country shortly before the enumeration.

⁶ Valuation in accordance with the value of a leisure hour is probably suitable only for proxy and personal surveys. In business surveys, respondents' time should be valued on an opportunity basis, i.e., in terms of the output that the respondent would generate during the time that she has to invest in responding to the survey—a tariff component in the Standard Cost Model (the "Dutch Model") that is applied, for example, in the UK to estimate the costs imposed on the business sector by government.

⁷ Under the reasonable assumption that C_T and μ are constants for all surveys.

⁸ Specifically, it does not imply an optimal choice of data collection mode or efficiency in organizing the fieldwork.

invested in producing questionnaires—from choosing the issues for investigation to developing and testing the questionnaire, in order to minimize the response burden.

21. Thus, it is possible to develop a survey-utility scale on the basis of the *ratio of time* that is needed to fill in the questionnaires. However, the benchmark for this scale—a survey that would serve as a basis for comparison—will be determined arbitrarily and by rules of thumb, organizational tradition, or trial and error. For any given benchmark survey, one may derive the utility scale an optimum number of attempts to interview for all other relevant surveys. Since attempts beyond the optimum number would not be worth economically, the optimal number would denote the cap on the number of attempts to interview in each survey, consistently and uniformly in all relevant surveys performed by the statistical organization.

22. The following algorithm makes the model easy to implement. First, we define one survey as a benchmark for determining the number of attempts to interview in other surveys, and for this benchmark we determine the maximum number of attempts, n_0^* . For any other survey k , the maximum number of attempts, n_k^* , is derived from n_0^* :

$$n_k^* = n_0^* \times \frac{T_k / (c_k h_k(n_0^*))}{T_0 / (c_0 h_0(n_0^*))} = n_0^* \times \frac{T_k}{T_0} \times \frac{h_0(n_0^*)}{h_k(n_0^*)} \times \frac{c_0}{c_k}, \quad (5)$$

where $h(n) = f(n)/S(n)$ is the hazard rate, valued at point n_0^* . Formula (5) states that the maximum number of attempts to interview in survey k will be greater insofar as completing its questionnaire is more time-consuming (i.e., the survey being more demanding), its enumeration hazard rate is lower, and its marginal enumeration costs are smaller—all relative to the benchmark survey.

23. In fact, the ratio of hazard rates in Formula (5) constitutes a “scale of measurement” for the transition from proxy surveys to personal surveys. Furthermore, Formula (5) may be applied to various sub-populations in one survey, with one sub-population serving as the benchmark. In this case, questionnaire length and marginal enumeration cost will be identical for the different sub-populations, since there is only one questionnaire and one collection method, and the maximum number of attempts to interview will be determined solely in view of the ratio of their hazard rates. Finally, the ratio of marginal enumeration costs as against the ratio of hazard rates in Formula (5) is a key factor in the transformation from face-to-face surveys to telephone surveys.

II. ANALYSIS OF THE ENUMERATION PROCESS’ DURATION

24. This section empirically investigates the relationship between the characteristics of locality, interviewer, and subjects, on the one hand, and the number of interview attempts that are needed to complete Part A of the HES questionnaire. First we present the response variable—the number of interview attempts and the results of the enumeration among Jews and Arabs, and the characteristics of the interviewers who worked on the 2004 Household Expenditure Survey.

25. Table 1 shows that there are significant differences between Jews and Arabs in the probability of successful enumeration. A larger proportion of Arabs than of Jews was interviewed upon the interviewer’s first attempt (53% vs. 34%, respectively). Shinar et al. (2005) suggest that the difference traces to a unique cultural characteristic of the Arab

population—the custom of hospitality, especially since the interviewers are Arab, as is usually the case in the HES. The reasons for failure to enumerate are different in each sector and the rate of non-contacts and refusals is significantly lower among Arabs than among Jews. Size of locality does not seem to affect the chances of enumeration from the first attempt; we observe no statistically significant difference between Jews and Arabs. The socioeconomic level of the locality also appears to have no perceptible effect on the enumeration rate. We test the effect of these factors below by means of a multivariate regression. The interviewer's personal characteristics, professional experience, and skills are known to affect the length and the outcome of the enumeration 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.⁹ Most HES interviewers are Jewish because Jews account for most of the sample and because Arab interviewers work mainly with Arab and Druze subjects and are not sent to localities that have Jewish majorities. Due to the sectoral distribution of the interviewers, one cannot determine whether the higher enumeration rate among Arabs originates in a higher response rate to the survey in the Arab sector or in superior performance of Arab interviewers relative to Jewish interviewers.¹⁰ Most interviewers are women. More than 70% are married; more than one-third have academic schooling. Arab interviewers are better educated and younger than Jewish interviewers; they also have less seniority as interviewers with CBS.

26. Now we estimate the marginal probability of enumeration among the survey sample. The factual enumeration rate as measured on the basis of the serial number of the attempt, according to the actual enumeration results (Figure 4), may be a biased estimate of the marginal probability of enumeration because the distribution is truncated from the right in some cases, i.e., attempts to interview “difficult” cases were halted after non-uniform numbers of attempts, at the coordinators' discretion. Thus, in these cases the factual enumeration results do not necessarily correspond to the results that would have been attained had the attempts continued.

27. Figure 5 shows the results of the a-parametric estimation of the marginal probability of enumeration using the Kaplan-Meier method. The figure shows clearly that both sectors evince a similar marginal probability of enumeration in attempts 2–7. 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 the duration of enumeration beyond seven attempts in the Jewish sector.

28. By performing a multivariate analysis of the cumulative probability of enumeration, using Formula (4), we may test the combined contribution of the characteristics of locality, interviewer, and subject. Our econometric model, which examines the effect of characteristics of locality, interviewer, and subject on the duration of enumeration by estimating a survival function, is different from the two econometric models that one encounters in the literature. In the traditional method (Groves and Couper, 1998; Stoop, 2004), researchers estimate a binary response (logit or probit) regression that focuses on the effects of various characteristics on

⁹ We have no information about “personal techniques” that interviewers use to persuade subjects to participate in the survey. Insofar as these factors affect interviewer effectiveness, an econometric model would reflect the influence of this factor as an interviewer fixed effect.

¹⁰ One may identify a separate effect of an interviewer and respondent's nationality only if Arab interviewers are given enumeration portions in the Jewish sector and vice versa. In practice, seven Arab interviewers worked on the Arab sector only and a few Jewish interviewers visited both sectors. Whether this variance suffices for separate identification of interviewer and subject nationality is an empirical issue that is examined below.

the probability of completing the interview of a subject in a given attempt. In fact, a binary regression “slices up” the continuity of the enumeration process: if the estimates of the explanatory variables change as the number of attempts to interview rises, it means that the respondents that are reached in high-order attempts have different traits than those interviewed in the first attempts. Nicoletti and Peracci (2005), who tested for dependency between the probability of contacting the subject and the probability of her cooperating with the interviewer in filling in the questionnaire, estimated a more sophisticated model: a censored bivariate probit model. They entered the number of attempts to interview as an explanatory variable in the likelihood-of-contact equation, and the sign of this variable, as expected, was negative (statistically significant). Unlike our econometric model, both of these estimation methods are susceptible to problems stemming from the truncation of the distribution of enumeration results, if the enumeration process is terminated before all cases are dealt with and their final status clarified. Furthermore, both methods disregard the implicit information in the serial number of each attempt to interview, i.e., they assume that the duration of enumeration has no effect on the probability of its success—an assumption that does not stand to reason.

29. Table 3 presents the results of survival-function estimation in four specifications. Model (a) includes only locality characteristics as regressors; model (b) adds interviewer characteristics; model (c) adds sixty additional dummy variables representing the interviewers’ fixed effects,¹¹ and model (d), without interviewers’ fixed effects, includes observed characteristics of the subjects from the completed questionnaires. The explained variable is the survival rate, i.e., the probability that the subject *will not be enumerated* in a given number of attempts. Therefore, a positive sign of the estimation denotes a variable that increases the number of attempts to interview, i.e., one that correlates negatively with the probability of enumerating the subject upon the first attempt.

30. The estimates of model (a) correspond to the picture elicited in Figure 4. Thus, more attempts are needed to enumerate Jews than to enumerate Arabs and the population of the three largest cities in Israel is “harder” to enumerate than that of smaller localities.

31. When interviewer characteristics are added, in model (b), the effect of subject’s nationality (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 with CBS were not found to have any effect. This last finding is somewhat surprising because one might expect duration of employment to reflect an accumulation of skill that would allow interviewers to convince subjects to cooperate in fewer attempts (Groves and Couper, 1998; Hox and de Leeuw, 2002). As a possible explanation to this finding, one may suggest that veteran interviewers are assigned “tougher” cases (e.g., more “problematic” neighborhoods) or have larger workloads. In fact, the average workload of an interviewer with less than two years’ seniority in the HES is 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 schooling helps to reduce the number of attempts (to a statistically significant extent among holders of masters’ degrees). Unmarried interviewers are not more effective than married

¹¹ 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.

ones; Jewish interviewers have to make more attempts than Arab interviewers to reach for the subjects in their workload.

32. The insertion of the interviewers' fixed effects in model (c) makes it impossible to pinpoint the effect of interviewer nationality.¹² Therefore, the effect of the Jewish localities becomes significant, as in model (a), although at a much lower intensity. In view of the fixed effects, the effect of schooling on reducing the number of attempts becomes stronger. Interviewer age and seniority with CBS 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 was not found to have any statistically significant effect in any model so far.

33. 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.¹³ 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.¹⁴ 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. This reduced the sample in model (d) to 5,850 households.

34. The results of model (d) show that when one controls for the effects of major cities respondents' characteristics, the socioeconomic level of the locality correlates negatively with the number of attempts to interview, i.e., enumeration is easier in wealthier localities. Fewer attempts are needed to enumerate households that are larger and have older and non-working 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's economic status may be reflected in the negative sign of the number-of-breadwinners variable.

35. The effect of interviewer characteristics is consistent with the findings discussed above: interviewers who hold masters' degrees are more effective than their less-educated peers and Arab interviewers make 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, all other factors being equal. However, interviewers with greater professional experience manage to enumerate 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 note that interviewers may tend to carry out more visits in places that are near their homes and along their routes of work.

¹²This is because Arabs worked in only the Arab sector and some Jewish interviewers worked in both sectors (in mixed-population towns).

¹³The time of visitation was not included among the explanatory variables even though the literature does customarily include it (e.g., Purdon et al., 1999, and Philippens et al., 2004). This is because enumeration in the Household Expenditure Survey takes place at 16:00–21:00, Sunday–Thursday, unless upon the first attempt the subject asks to be visited in the morning or early afternoon.

¹⁴The characteristics of those who drop out of the survey when asked to fill in the expenditure diary are a topic for separate research.

III. LIMITING THE NUMBER OF INTERVIEW ATTEMPTS AND HOW LIMITATION AFFECTS THE SURVEY ESTIMATES

36. In this section we perform three simulations in which the number of attempts to interview is limited differentially for subjects in Jewish and Arab sectors. The Arab sector will serve as a benchmark population for determining the cap on the number of interview attempts. In regard to this benchmark sector, we set (arbitrary) two alternative caps: three and four attempts. On the basis of Formula (5), the maximum number of attempts in the Jewish sector is set in accordance with the ratio of hazard rates in the two sectors that were estimated in the previous section (Figure 5):

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

37. Thus, if the cap on the number of attempts in the Arab sector is set at three, there will be four attempts at the most in the Jewish sector, and if a maximum of four attempts is made in the Arab sector, the respective cap in the Jewish sector should be seven. To analyze the sensitivity of the cap in the Jewish sector (in which the maximum leaps from four to seven), we perform a simulation of an in-between situation with a maximum of four attempts to interview Arab subjects and five for Jewish subjects.

38. We then compare the results of each simulation with the results of the actual enumeration. Thus, 6,132 of the 7,782 households in the sample of the 2004 Household Expenditure Survey completed the survey.¹⁵ The number of attempts to interview at the phase of completing Part A of the questionnaire was not limited in a uniform way; instead, the decision about whether to continue attempting the subject or to desist was made on a case-by-case basis; it ranged from 1 to 22 in the Jewish sector and 1 to 13 in the Arab sector, while the average number of attempts per subject was 2.67 and 1.93, respectively (Table 4).

39. The three simulations reflect a situation in which the number of attempts was limited separately for each sector. Due to the imposition of the cap, some cases that were actually identified as ineligible for interviewing would be classified as non-contacts and some cases that actually completed the interview but after more attempts than the cap would allow would not be enumerated. The stricter the limit, the fewer cases were processed and the fewer subjects were fully enumerated. Thus, the proportion of fully enumerated subjects in Simulation A, in which the limit was strictest, was 77.2%, whereas in Simulation C, which had the most liberal limit, it was 85.8%—as against 88.3% in the actual survey, which, as said above, imposed no general limit on the number of attempts to interview (Table 4).

40. Obviously, limiting the number of attempts saves interviewers' time (and organization's budget). The saving ranged from 14% in Simulation A to 3.5% in Simulation C. One may "invest" the number of attempts saved by limiting the efforts to interview the hard-to-reach respondents by increasing the size of the sample. In Table 5, we calculate the added increment to the final sample (the number of completed interviews) that would come about if the survey sample were expanded at the price of forgoing "difficult" cases.

¹⁵ That is, they completed Part A of the questionnaire, the expenditure diary, 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.

41. The situation in which cases that require too many attempts to interview are forgone and replaced by sampling new cases from the survey frame is akin to “refreshment” in the sense of Hirano et al. (2001)’s work. In the problem that Hirano et al. discuss, sample “refreshment” is meant to mitigate the attrition bias in a longitudinal (panel) survey, unless the attrition process sustains the premise of missing-at-random. In regard to our problem, the purpose of “refreshment” is to alleviate the non-response bias due to the omission of the hard-to-reach part of the sample in a cross-section survey. Let us note that hard-to-reaches are largely of the non-contact type (their share will even rise after putting a cap on the number of attempts to reach)—see Table 1.16. Let us also recall that their characteristics closely resemble those of populations that are strongly predisposed to attrition in panel surveys—households with high geographic mobility, composed of individuals or childless couples, living in rented dwellings, or weakly attached to the labor market.¹⁷ Thus, sample-refreshment, as a solution to attrition in a panel survey, may also solve the problem of an increasing proportion of non-contacts in a cross-section survey. In our simulation, where the increment added by means of sample-refreshment is synthetic only—reproducing the traits of a sample that is limited in number of interview attempts—we have no way of showing empirically the improvement in estimate precision that is expected according to Hirano et al. (2001). However, we should remember that the estimates of the survey under the simulation terms, as presented below, are from a survey sample that is less than complete (89% in Simulation A, 94% in Simulation B, 98% in Simulation C). Therefore, increasing the sample ought anyway to improve the estimates precision.

42. In Table 5, this sample-refreshment addition is assumed to be enumerated in accordance with the three chosen simulation scenarios. Some of these subjects were fully enumerated—at the same proportion of fully enumerated subjects that was attained in a given simulation for the original sample, as shown in Table 4 (e.g. 75.7% in Simulation A in the Jewish sector). It turns out that the expected number of fully enumerated subjects, including the sample-refreshment addition, that was attained due to the saving on attempts to interview surpassed the number of subjects who were fully enumerated in the actual survey. The increases were 4% in Simulation A, 3% in Simulation B, and 1% in Simulation C. This result is intuitive because hard-to-reach marginal cases, which require more-than-average attempts to interview, would be replaced with average cases, whose enumeration is 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 (Figure 6).

43. Since any effective cap on the number of attempts will lower the response rate, the main question is whether the survey estimates will be biased due to the uniqueness of the characteristics of hard-to-reach subjects. To answer this question, we put three samples, obtained as a result of limiting the number of interview attempts in the aforementioned simulations, through the standard process of constructing weights, as accustomed in the HES.

44. The set of weights coefficients is constructed in a multiphase 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

¹⁶ Abraham et al. (2006) report in regard to the American Time Use Survey that 60% of non-respondents are non-contacts.

¹⁷ Nicoletti and Peracchi (2005) and Stoop (2004) found these factors capable of explaining low contact probability. Fitzgerald et al. (1998) report the same traits as characteristic of dropouts from PSID panel surveys.

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.) and seasonal adjustment of the enumeration period. The characteristics of individuals are sex and age group, by geographic regions. (For details, see Central Bureau of Statistics, 2006.)

45. The construct is such that the estimates of the distribution of these calibrated characteristics should not be affected by a slight decline in the response rate if the response rate remains high, e.g., the 76%–85% rates in our simulations. 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 hard-to-reach part of the sample by capping 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 three simulations are 11%, 6%, and 2% smaller than the final sample of the survey. On the other hand, if the populations that are not enumerated due to the limitation of attempts are much different from those who remained in the sample, and then the estimates may be more precise but biased. The buildup of the sample with a “refreshment” addition is meant to reduce, if not eliminate, the first of these influences. Its ability to counteract the second influence, however, is definitely limited.

46. Table 6 presents estimates for a range of household characteristics across the full survey sample and the three 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. Naturally, the more interview attempts are allowed, the more closely the truncated samples will approximate the actual sample. Therefore, we focus our discussion mainly on Simulation A, which underscores the influence of sample truncation on the quality of the estimates. 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 interview attempts needed for enumeration (Table 3).

47. Before we analyze the weighted estimates of the survey variables, we should note that the variation coefficient of the weights declined, i.e., their variance widened. General observation of the estimates of the traits of the survey population shows that the estimates for the Arab sector are less robust in all simulations, although the differences relative to the actual sample are rarely statistically significant. Since the sample in Simulation C is only 3% smaller than the actual sample, one may infer that lack of robustness in the estimates (a typical feature of all the 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.

48. 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 all simulations) and pension receipts (in Simulations A and B). These populations are known for their less-than-complete coverage in the HES. 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.¹⁸ Since administrative sources 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.

49. 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). Increasing the sample by cutting back on attempts to interview may make the estimates more precise and reduce the bias that has come about.

IV. CONCLUSION AND DISCUSSION

50. Face-to-face interviewing is the most expensive method that a statistical organization can use to gather information. 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. Even if they are, in most countries the major surveys, such as Labour Force surveys and Household Expenditure or Income surveys, continue to be performed by means of fieldwork. Therefore, the streamlining of fieldwork remains a significant issue.

51. Some may claim that the remuneration of interviewers by output (number of subjects enumerated) absolves the organization of the need to limit the number of interview attempts because, practically speaking, it shifts responsibility for deciding how many attempts are worthwhile onto the interviewer, who is supposed to strike a balance between the circumstances of each subject in her workload and the probability of success in view of a predetermined level of remuneration. This solution, however, is inapplicable when the interviewers are civil servants, who are not customarily paid on a piecework basis. When piecework remuneration of interviewers is inapplicable, the organization probably comes out at a disadvantage: the range of different considerations among interviewers and the profusion of contacts with subjects¹⁹ may impair the quality of the data and generate public resistance to data-gathering operations. Accordingly, the question of setting a maximum number of attempts to interview or contact subjects should be answered by the organization, irrespective of how it remunerates its interviewers.

52. Every organization solves the matter in its own way. Some national organizations impose limits (three attempts in the Netherlands, four attempts in the UK) and some (e.g., Israel and Canada) do not. In this paper, we present a cost/benefit model in which the optimum number of attempts or requests is determined at the intersection of the marginal cost of attempts and the expected utility of the information. Beyond this number of attempts, the enumeration effort is not economically justified.

53. The main findings of the multivariate regression analysis point to factors that influence the number of attempts to interview. Some of these variables are known *ex ante* (before the enumeration begins). Thus, locality size, subject's religion, and interviewer characteristics

¹⁸ According to the income-tax records, there were 307,000 active self-employed persons in Israel in 2004 (excluding corporate executives who supposedly defined themselves in the survey as wage-earners), as against 256,000 according to the estimate in the actual sample.

¹⁹ This result is known from the economic literature that discusses the advisability of privatizing tax collection. (For a survey, see Slemrod and Yitzhaki, 2002.) A tax farmer, who operates to maximize his profit, will exert him/herself more than society needs because she does not take into account the excess burden of the subjects response to h/her requests, since this is accounted for as a social cost.

were found to affect the chances of successful enumeration. Other factors that become known *ex post* (after enumeration), including household size, age of head of household, and non-working status of head of household, were also found to have an influence. These findings show that the probability of enumeration varies among subjects and characteristics, suggesting the possibility of setting different caps on the number of interview attempts in accordance with these variables.

54. To demonstrate the point, we applied the model in three simulations on the basis of the 2004 HES. 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 and were compared with corresponding estimates from the actual sample, which was surveyed with no effective limit on the number of interview attempts.

55. Due to the cap on interview attempts, the response rate declined at increments ranging from 11 percentage points (in the simulation that imposed a cap of three interview attempts in the Arab sector and four attempts in the Jewish sector) to 2 percentage points (in the simulation that imposed a cap of four and seven attempts, respectively). As the findings showed, the decline in response rate 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 cap 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, who are also known for underreporting in the full survey, in which the number of interview attempts is not limited. 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.

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

57. While considerable savings may be achieved by making fewer interview attempts, the effect of such an action on the management of fieldwork should be considered, especially if the limitation of interview attempts is to be differentiated on the basis of subjects' and/or interviewers' characteristics. The message that the interviewer hears today is that she must make every attempt to maximize the rate of enumeration. Interviewers are paid on the basis of hours worked, it being assumed that this will prevent them from "scrimping" on repeat visits and allow them to deal with cases that are hard to reach or interview. Concurrently, their enumeration percentages are susceptible to powerful pressure and controls, coupled with a persistent demand to raise the enumeration rates. If the organization limits the number of interview attempts *ab initio*, it may send the interviewer a signal that the goal is not to obtain the highest enumeration rates; this may impair her motivation. Therefore, to apply the model it would be necessary to introduce changes in interviewer training and control and evaluation methods.

58. The practical implementation of the model also depends on the level of resources that the interviewer has to invest in each attempt. The model assumes that all interview attempts are equal in cost and quality. In practice, however, interviewers tend to make more attempts when they do not need to overexert themselves in so doing, e.g., by making lengthy trips to subjects' homes. 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 special quantities of time or effort and where the subject is on the interviewer's route in any case. The findings about interviewer characteristics that tend to reduce the number of attempts may lead to conclusions about revising the process that is 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 a larger proportion of well-schooled people for this job. Furthermore, there are requirements that impede the identification of appropriate interviewers, including fluency in languages other than Hebrew (Arabic, Russian). If so, it may be appropriate to ask whether interviewer training can be improved in order to enhance the interviewers' abilities and motivation. Training activities may alleviate various deficiencies in the enumeration process; it is not clear that such activities are being offered adequately.

59. Additional difficulties of administrative nature arise when a differential number of attempts to interview, e.g. in accordance with sample composition, is set. It is plausible that the model should not be applied at the level of the individual subject but rather for relatively large population groups, e.g., the Arab and Jewish sectors, as we did in the simulation. An organization that decides to establish differential caps on the number of attempts to interview will have to take at least one of the following two actions: training interviewers to make decisions about investing resources in accordance with sampling unit characteristics that are disclosed to them in advance, or assuring appropriate instructions from the coordinators for each sampling unit. Each of these actions entails a new array of control and supervision.

60. Despite the difficulties that arise when the number of interview attempts in fieldwork is limited, it should be remembered that such a limitation may help the organization achieve a meaningful saving of resources that can be reinvested in other measures that the surveying work requires, e.g., increasing the sample size or improving performance by means of instructional and training activities. Therefore, other surveys should be examined in order to corroborate the findings, so that the policy on enhancing the efficiency of fieldwork, and the requisite deployment for such a policy, will be compatible among different surveys.

BIBLIOGRAPHY

- Abraham, K. G., Maitland, A., Bianchi, S. M. (2006). Non-response in the American Time Use Survey: who is missing from the data and how much does it matter? *National Bureau of Economic Research, Technical Working Paper Series*, No. 328.
- Central Bureau of Statistics (2006). *Household Expenditure Survey 2004, General Summary*, Special Publication 1261.
- Fitzgerald, J., Gottschalk, P., Moffitt, R. (1998). An analysis of sample attrition in panel data: the Michigan Panel Study of Income Dynamics. *Journal of Human Resources* 33, 251–299.
- 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 (5), 646-675.
- Groves, R. M., Couper, M. P. (1998). *Nonresponse in Household Interview Surveys*, Ch. 7, New York: John Wiley and Sons.

- Hirano, K., Imbens, G., Ridder, G., Rubin, D. B. (2001). Combining panel data sets with attrition and refreshment samples. *Econometrica* 69 (6), 1645–1659.
- Hox, J., de Leeuw, E. (2002). The influence of interviewers' attitude and behavior on household survey nonresponse: an international comparison. In: *Survey Nonresponse*, edited by R. M. Groves, D. A. Dillman, J. L. Eltinge, and R. J. A. Little, 103–120. New York: John Wiley and Sons.
- Lynn, P., Clarke, P., Martin, J., Sturgis, P. (2002). The effects of extended interviewer efforts on nonresponse bias. In: *Survey Nonresponse*, edited by R. M. Groves, D. A. Dillman, J. L. Eltinge, and R. J. A. Little, 135–148. New York: John Wiley and Sons.
- Nicoletti, C., Peracchi, F. (2005). Survey response and survey characteristics: Microlevel evidence from the European Community Household Panel. *Journal of the Royal Statistical Society* 168 (4), 763–781.
- Lebrasseur, D., Dian, 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., Clarke, P. (2002). Separating refusal bias and non-contact bias: Evidence from UK national surveys. *The Statistician* 51 (3), 319–333.
- Philippens, M., Billiet, J., Loosveldt, G., Stoop, I., and Koch, A. Work package 7. Data-based quality assessment in the ESS, Part 2. Non-response and fieldwork efforts in the ESS: Results from the analysis of call record data.
http://naticent02.uuhost.uk.uu.net/methodology/nonresponse_fieldwork_efforts.doc
- Purdon, S., Campanelli, P., Sturgis, P. (1999). Interviewers' calling strategies on face-to-face interview surveys. *Journal of Official Statistics* 15 (2), 199–216.
- SAS/STAT User's Guide. SAS Institute Inc., Cary, NC, USA.
- Shinar, S., Oren, O., 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. In: *Handbook of Public Economics*, edited by A. J. Auerbach and M. Feldstein, edition 1, volume 3, 1423–1470. Elsevier.
- Stoop, I. (2004). Surveying nonrespondents. *Field Methods*, vol. 16 (1), 23–54.
- Sturgis, P., Smith, P., 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., Sprachman, S. (2003). Costs and benefits of improving response rates for a hard-to-reach population. *Public Opinion Quarterly* 67, 126–138.

Figure 1: Marginal Cost, Expected Marginal Benefit, and Optimum Number of Attempts to Interview

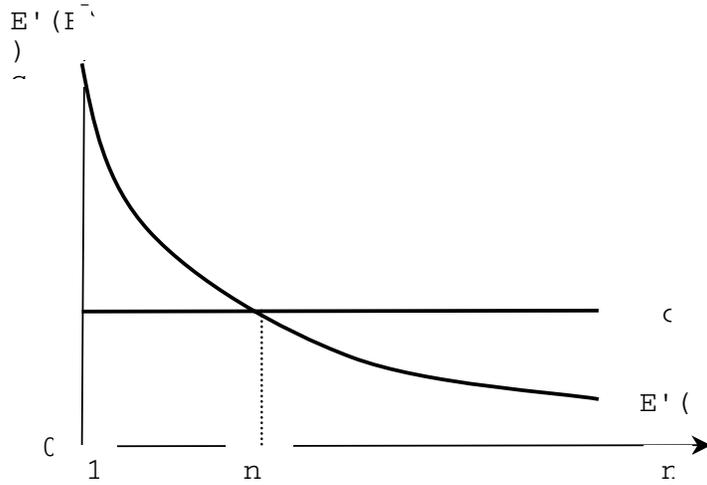


Figure 2: Marginal Cost, Expected Marginal Benefit, and Optimum Number of Attempts to Interview, for Two Populations Differentiated in Probability of Enumeration

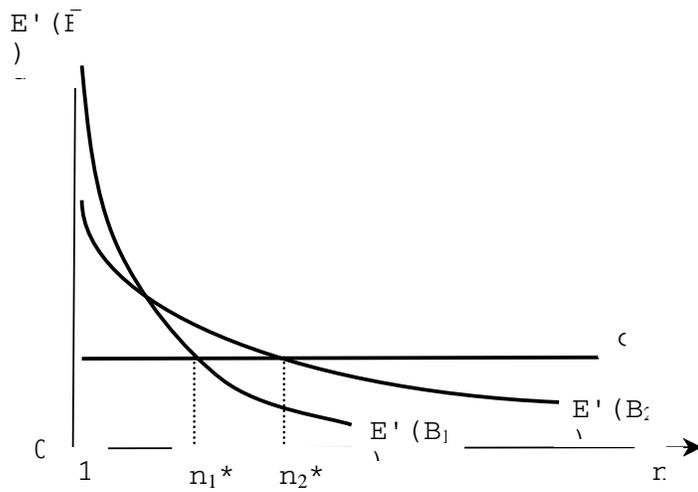


Figure 3: Marginal Cost, Expected Marginal Benefit, and Optimum Number of Attempts to Interview, for Proxy and Personal Survey

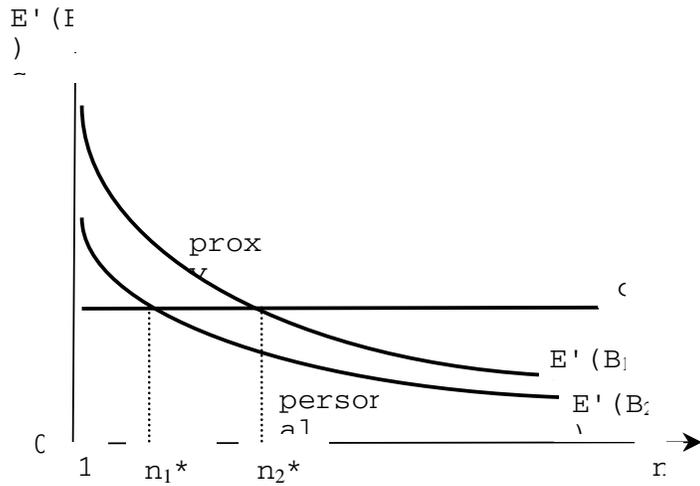
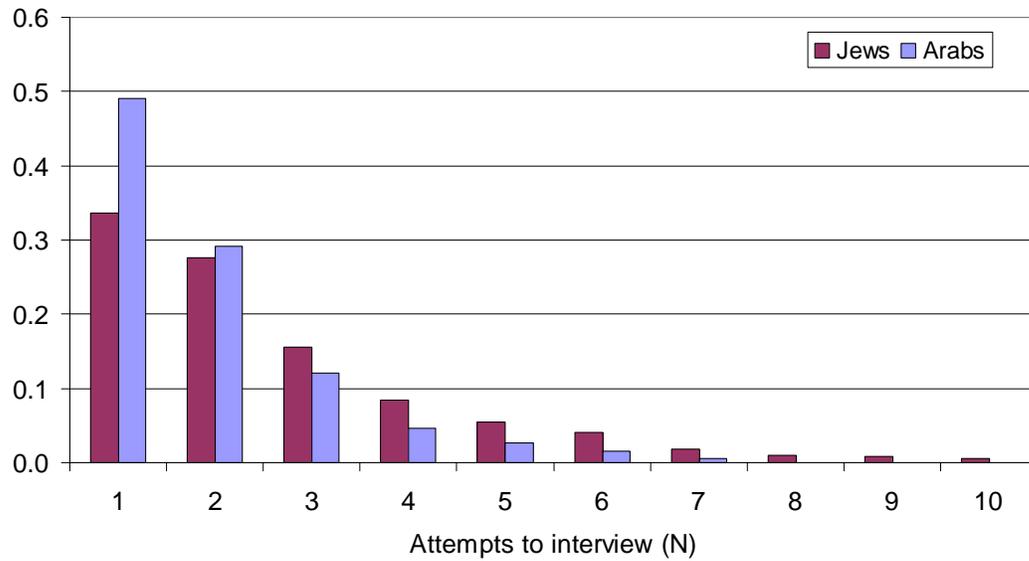
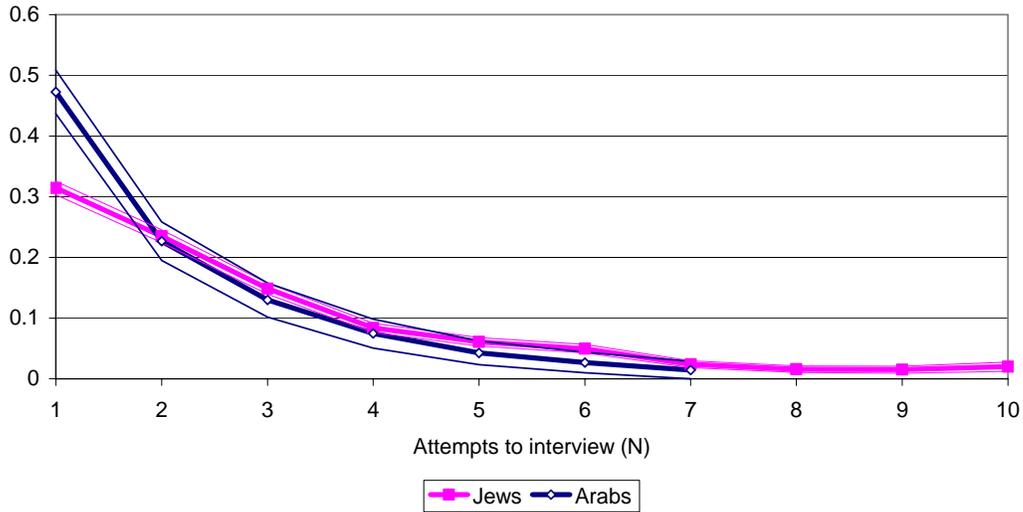


Figure 4: Distribution of Number of Interview Attempts



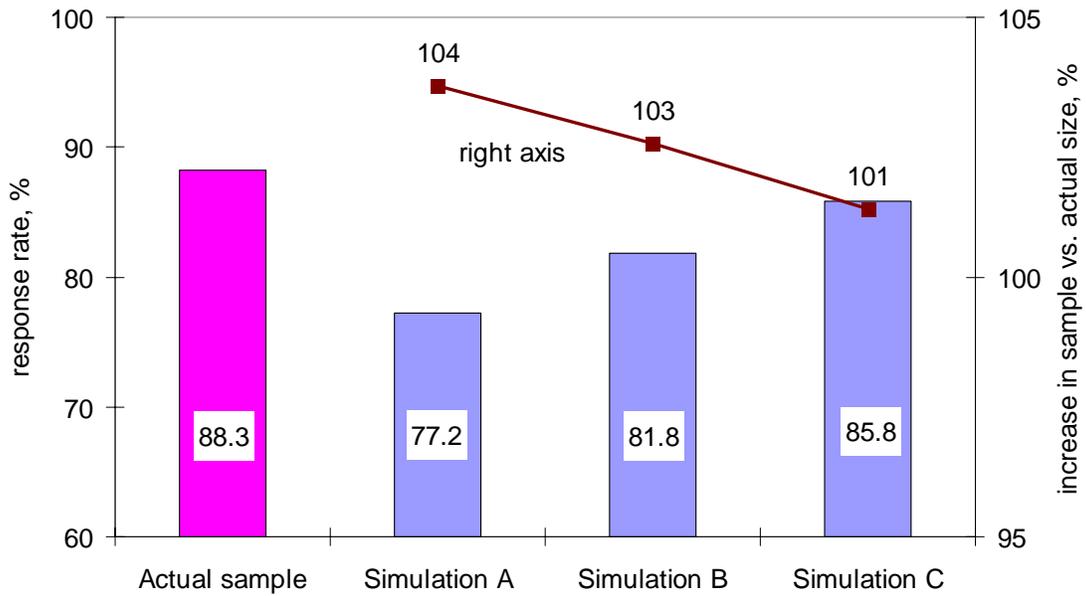
Source: 2004 Household Expenditure Survey.

Figure 5: Marginal Probability of Enumeration, by Number of Attempts to Interview, by Population Group (with 95% confidence interval)



Source: 2004 Household Expenditure Survey, authors' computations.

Figure 6: Response Rate and Sample Size, if Increased on Account of Forgone Interview Attempts



Source: 2004 Household Expenditure Survey, authors' computations from Table 5.

Table 1: Enumeration Results by Population Group, Locality Size, and Locality Socioeconomic Level¹ (percent)

	Total	Jews	Arabs
Total sample	100.0	100.0	100.0
Non-eligible	10.8	10.7	11.8
Refusal	4.2	4.4	2.1
Non-contacts	4.8	5.2	1.1
Partial completion of questionnaire	1.1	1.0	1.9
Fully enumerated (excl. non-eligible 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>
Rishon Leziyyon	90.8	90.8	
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 Household Expenditure Survey, authors' computations

(1) “Jews” includes others; “Arabs” includes Moslems, Christians, and Druze. Locality size is as of 2004. Socioeconomic level is based on the CBS 2001 socioeconomic index, which has a scale of 1 (low) to 10 (high).<> = <20 observations in cell.

Table 2: Enumeration Results and Characteristics of Interviewers in Household Expenditure Survey, by Interviewer's Nationality¹ (percent unless otherwise noted)

	All Interviewers	Jews	Arabs
Number of Interviewers	66	59	7
Rate of non-eligible	10.9	11.1	8.8
Rate of refusal	4.1	4.4	1.4
Rate of non-contacts	9.9	10.5	3.8
Rate of fully enumerated (excl. non-eligible 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 Household Expenditure Survey, authors' computations.

(1) In calculating the enumeration results, 206 observations omitted because their records lacked interviewer's details.

Table 3: Estimates of Survival Function of Interview Attempts to Complete Questionnaire Part A¹

Independent variable	(a)	(b)	(c)	(d)
Jewish locality ²	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)**
Rishon Leziyyon	0.085 (0.054)	0.084 (0.054)	-0.043 (0.137)	0.026 (0.050)
Locality socioeconomic level ³	-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)
Woman		-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 CBS 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 ⁴	None

Observations (N)	7576	7411	7348	5850
Log-likelihood	-8865.2	-8656.9	-8422.3	-6232.0

Source: 2004 Household Expenditure Survey, authors' computations

- (1) Standard deviations are in parentheses. * 0.05 significant estimate, ** 0.1 significant estimate. Omitted 206 observations for which interviewer details were not recorded.
- (2) Nationality of locality is defined at the level of statistical region by the religion of a majority of its inhabitants according to the 1995 census. "Jews" includes non-Arabs; "Arabs" includes Moslems, Christians, and Druze.
- (3) Socioeconomic level is based on the 2001 socioeconomic index, on a scale of 1 (low) to 10 (high).
- (4) Of 60 fixed effects, 40 were significant at a 0.05 level.

Table 4: Simulation of Limitation on Number of Interview Attempts for Completion of Questionnaire Part A, by Population Group¹

	Total	Jews	Arabs
Total households in sample, gross	7782	6763	1019
Actual enumeration			
Maximum number of attempts		22	13
Non-eligible	834	746	88
Fully enumerated	6132	5241	891
Response rate, percent, excl. non-eligible cases	88.3	87.1	95.7
Total attempts to interview	20051	18086	1965
Average attempts per subject	2.58	2.67	1.93
Simulation A			
Maximum number of attempts		4	3
Non-eligible	676	594	82
Fully enumerated	5488	4673	815
Response rate, percent, excl. non-eligible cases	77.2	75.7	87.0
Total attempts to interview	17260	15472	1788
Average attempts per subject	2.22	2.29	1.75
Simulation B			
Maximum number of attempts		5	4
Non-eligible	730	645	85
Fully enumerated	5771	4917	854
Response rate, percent, excl. non-eligible cases	81.8	80.4	91.4
Total attempts to interview	18357	16477	1880
Average attempts per subject	2.36	2.44	1.84
Simulation C			
Maximum number of attempts		7	4
Non-eligible	787	702	85
Fully enumerated	6002	5138	864
Response rate, percent, excl. non-eligible cases	85.8	84.8	92.5
Total attempts to interview	19371	17472	1899
Average attempts per subject	2.49	2.58	1.86

Source: 2004 Household Expenditure Survey, authors' computations

- (1) Nationality of fully enumerated households is defined by religion of head of household. For those not fully enumerated and for which head of household's religion is not known—by nationality of locality (see note 2 to Table 3). Since the definition of subject's nationality in this table is different from that in Table 1, the proportion of those fully enumerated (out from eligible) is slightly different than that shown in Table 1.

Table 5: Simulation of Limitation on Number of Interview Attempts for Completion of Questionnaire Part A, by Nationality of Subject¹

	Total	Jews	Arabs
Actual enumeration			
1. Fully enumerated	6132	5241	891
Simulation A			
2.1 Fully enumerated	5488	4673	815
2.2 Total interviewer attempts saved ²	2791	2614	177
2.3 Possible increase in survey sample ³	1243	1143	101
2.4 Expected addition of completed interviews due to increase in sample ⁴	870	790	81
2.5 Total expectation of completion of interview [(2.1)+(2.4)]	6358	5463	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			
3.1 Completion of interview	5771	4917	854
3.2 Total interviewer attempts saved ²	1694	1609	85
3.3 Possible increase in survey sample ³	706	660	46
3.4 Expected addition of completed interviews due to increase in sample ⁴	519	480	39
3.5 Total expectation of completion of interview [(3.1)+(3.4)]	6290	5397	893
3.6 Ratio of expected completion of interview to actual completion of interview $[(3.5)/(1)]$	1.03	1.03	1.00
Simulation C			
4.1 Completion of interview	6002	5138	864
4.2 Total interviewer attempts saved ²	680	614	66
4.3 Possible increase in survey sample ³	273	238	35
4.4 Expected addition of completed interviews due to increase in sample ⁴	211	181	30
4.5 Total expectation of completion of interview [(4.1)+(4.4)]	6213	5319	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

- (1) Nationality of subject: nationality of fully enumerated households is defined by religion of head of household. For those not fully enumerated and for which head of household's religion is not known—by nationality of locality (see note 2 to Table 3).
- (2) The difference between total attempts to interview in actuality and total attempts made in a given simulation (Table 4).
- (3) The increase in the survey sample that could be afforded due to saving of field-staff time with the cap on interview attempts. The added increment is total attempts saved divided by average attempts per subject in a given simulation (Table 4).
- (4) The expectation of completed interview due to the increase in the sample is the added increment to the survey sample multiplied by the rate of completion of interview among members of the frame in a given simulation.

**Table 6: Estimates of Demographic and Economic Characteristics from the 2004 Household Expenditure Survey—
Actual Sample and Three Truncated Samples, by Nationality¹ of Head of Household* (thousands)**

	Actual sample			Simulation A			Simulation B			Simulation C		
	Total	Jews	Arabs	Total	Jews	Arabs	Total	Jews	Arabs	Total	Jews	Arabs
Households in sample	6132	5241	891	5488	4673	815	5771	4917	854	6002	5138	864
Household composition, non-weighted sample (%)												
1 person		18.5	7.1		17.6	6.6		17.7	6.8		18.2	7.2
2 persons		26.5	13.6		26.3	13.3		26.4	13.6		26.4	13.5
3 persons		16.8	11.2		16.9	11.7		16.9	11.1		16.8	11.1
4 persons		17.6	18.6		18.0	18.8		18.0	18.6		17.8	18.5
5+ persons		20.6	49.5		21.2	49.6		21.0	49.9		20.8	49.7
Avg. weight	318.4	315.7	329.2	351.3	350.1	355.6	334.7	333.2	340.7	324.3	320.6	339.2
Weight c.v.	2.066	2.348	1.535	2.007	2.226	1.537	2.034	2.301	1.511	2.040	2.325	1.515
Survey weighted population	6494.0	5137.8	1356.2	6494.4	5149.9	1344.5	6494.4	5143.6	1350.8	6494.4	5139.5	1354.9
Age groups												
0–17	2146.8	1544.0	602.7	2146.7	1550.6	596.1	2146.8	1546.5	600.3	2146.7	1545.2	601.5
18–24	725.5	556.6	168.9	726.4	556.6	169.8	725.6	556.4	169.2	725.7	555.8	169.9
25–34	1004.0	796.8	207.2	1004.0	800.0	204.0	1004.5	799.1	205.3	1004.5	798.2	206.3
35–44	769.6	602.2	167.4	768.5	604.2	164.3	769.1	602.9	166.2	769.6	602.6	167.0
45–54	722.5	621.3	101.3	723.0	621.8	101.2	723.1	621.8	101.3	722.6	621.0	101.6
55–69	698.1	621.3	76.9	698.3	620.7	77.6	698.3	620.3	78.0	698.4	620.9	77.6
70+	427.4	395.7	31.7	427.6	396.0	31.6	427.1	396.6	30.4	427.0	395.5	31.0
Marital status												
Married	2721.9	2224.4	497.6	2720.7	2227.3	493.4	2720.2	2224.9	495.3	2719.9	2223.4	496.5
Single	1067.1	862.7	204.3	1074.9	871.0	204.0	2071.7	866.8	204.9	1068.8	863.3	205.6
Divorced	211.8	203.9	7.8	207.3	199.5	7.8	208.4	200.5	7.9	212.9	205.0	7.9
Widowed	295.4	258.8	36.6	295.1	258.1	37.0	295.9	260.2	35.7	295.0	258.8	36.1

1989+ immigrants	945.1	875.5	69.6	944.3	878.0	66.2	949.7	880.4	69.3	946.3	875.6	70.7
Education												
Secondary	2004.0	1636.2	367.8	2010.5	1644.6	365.8	2013.0	1645.5	367.5	2007.1	1638.6	368.5
Post-secondary	705.5	642.0	63.4	694.1	633.4	60.7	699.4	637.1	62.3	704.1	640.8	63.3
non-academic												
Academic	1212.8	1111.8	101.0	1222.2	1121.4	100.7	1216.2	1114.6	101.6	1211.5	1110.5	101.0
Labor-force characteristics and sources of income												
Worked (in three months preceding survey)	2294.1	1986.4	307.7	2297.5	1995.8	301.7	2296.2	1991.1	305.1	2295.0	1987.4	307.6
Employee	2038.2	1772.7	265.5	2052.3	1787.6	264.7	2045.7	1779.9	265.8	2040.7	1773.4	267.3
Self-employed	255.9	213.7	42.2	245.2	208.2	37.0	250.5	211.2	39.3	254.3	214.0	40.3
Employees, worked <35 hours per week	588.7	522.5	66.2	591.1	525.7	65.5	591.5	527.2	64.3	591.2	525.8	65.4
36–45 hours per week	745.7	657.6	88.1	754.6	666.2	88.4	751.3	661.0	90.3	745.6	656.0	89.6
46+ hours per week	703.9	592.7	110.9	706.6	595.8	110.9	702.9	591.8	111.2	703.8	591.6	112.3
Child-allowance recipients	896.8	702.2	194.6	894.4	700.8	193.6	895.3	701.8	193.5	897.2	702.7	194.6
Disability-benefit recipients	173.8	140.9	32.9	177.5	142.8	34.6	173.2	139.5	33.7	173.8	140.8	33.0
Unemployment-compensation recipients	56.6	46.6	10.0	56.1	46.0	10.2	57.1	46.6	10.5	57.0	46.6	10.4
Income-maintenance recipients	130.4	86.5	43.9	134.3	89.2	45.0	130.9	86.2	44.7	131.2	86.7	44.5
Net income, employee and self-employed, NIS/month	5580	5739	4554	5581	5741	4518	5577	5735	4547	5584	5744	4551
Pension income in Israel, NIS/month	4556	4602	3575	4465	4504	3658	4484	4528	3567	4560	4608	3564

Total expenditure per capita, NIS/month	2427	2633	1648	2432	2641	1633	2432	2639	1644	2429	2637	1640
Total consumption expenditure, NIS/month	10442	10627	9385	10458	10653	9330	10461	10647	9391	10450	10642	9352
Thereof:												
Food (excl. fruit and vegetables)	1377	1331	1636	1383	1338	1644	1378	1332	1642	1377	1332	1638
Fruit and vegetables	344	326	446	345	327	448	345	327	449	344	326	447
Housing	2343	2459	1681	2335	2448	1684	2336	2450	1685	2342	2458	1682
Home and household upkeep	1052	1083	872	1055	1086	874	1052	1083	874	1052	1084	872
Home furniture and appliances	588	564	743	582	551	782	586	560	757	586	561	749
Clothing and footwear	588	539	815	590	540	824	590	540	825	589	539	821
Healthcare	568	574	531	570	577	528	571	577	532	569	576	525
Education, culture, and entertainment	1433	1528	887	1444	1541	877	1439	1538	868	1434	1534	863
Transport and communications	2141	2204	1772	2134	2206	1711	2145	2207	1780	2144	2206	1775
Other goods and services	565	548	656	567	556	629	569	557	637	564	551	634

Source: 2004 Household Expenditure Survey, authors' computations

* Simulation A: up to 4 attempts to interview Jews, up to 3 attempts to interview Arabs; * Simulation B: up to 5 attempts to interview Jews, up to 4 attempts to interview Arabs; * Simulation C: 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.

(1) Nationality of head of household—"Jews" includes non-Arabs.