

Johansson, Fredrik; Klevmarken, Anders

Working Paper

Explaining the size and nature of response in a survey on health status and economic standard

Working Paper, No. 2006:2

Provided in Cooperation with:

Department of Economics, Uppsala University

Suggested Citation: Johansson, Fredrik; Klevmarken, Anders (2006) : Explaining the size and nature of response in a survey on health status and economic standard, Working Paper, No. 2006:2, Uppsala University, Department of Economics, Uppsala, <https://nbn-resolving.de/urn:nbn:se:uu:diva-83065>

This Version is available at:

<https://hdl.handle.net/10419/82752>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



UPPSALA
UNIVERSITET

Working Paper 2006:2

Department of Economics

Explaining the size and nature of response in a survey on health status and economic standard

Fredrik Johansson and Anders Klevmarken

Department of Economics
Uppsala University
P.O. Box 513
SE-751 20 Uppsala
Sweden
Fax: +46 18 471 14 78

Working paper 2006:2
Januari 2006
ISSN 0284-2904

EXPLAINING THE SIZE AND NATURE OF RESPONSE IN A SURVEY
ON HEALTH STATUS AND ECONOMIC STANDARD

FREDRIK JOHANSSON AND ANDERS KLEVMARKEN

Papers in the Working Paper Series are published
on internet in PDF formats.

Download from <http://www.nek.uu.se>
or from S-WoPEC <http://swopec.hhs.se/uunewp/>

Explaining the size and nature of response in a survey on health status and economic standard*

Fredrik Johansson[†] and Anders Klevmarken[‡]

Department of Economics
Uppsala University

January 5, 2006

Abstract

Using rich register data to analyze response behavior in a survey on health and economic standard, a model to explain contact and participation probabilities is estimated. A main result is that both probabilities are lower among respondents out of the labor market, who are immigrants and on benefits.

JEL-classification: C34, C42, J14.

Keywords: Nonresponse, Response process

* The authors are grateful for suggestions on a previous version from Arie Kapteyn, Susann Rohwedder and Jelmer Yeb Ypma. Thanks also to seminar participants at the Department of Information Science, Division of Statistics, Uppsala University. This paper is part of an NIA (R03AG21780) and FAS (2001-2830) funded project (Comparison of Survey and Register Data: The Swedish Case) in collaboration with Arie Kapteyn and Susann Rohwedder (RAND).

[†] Department of Economics, Uppsala University, P.O. Box 513, SE-75120 UPPSALA, Sweden. Phone: 46-18-471 16 36. Email: Fredrik.Johansson@nek.uu.se.

[‡] Department of Economics, Uppsala University, P.O. Box 513, SE-75120 UPPSALA, Sweden. Phone: 46-18-471 23 86. Email: Anders.Klevmarken@nek.uu.se.

1. Introduction and motivation

Nonresponse is probably the most severe problem in survey research. Today it is not unusual to find surveys with a response rate around or even below 50 percent. It is obvious that this high nonresponse will not only decrease sample size and correspondingly increase variances of estimates from these data, but the results might also become biased if response is selective. There is a large literature on methods to compensate for nonresponse ranging from calibration methods including standard post stratification, imputations, and to more sophisticated model based methods. The key to a successful compensation is to understand the causes of nonresponse. This is also important because of its relevance to survey design, where resources have to be allocated between the possibly conflicting goals of increasing the precision of estimates and reducing nonresponse biases.

There is an increasing literature on the causes of nonresponse with more or less successful attempts to build models explaining response behavior. These attempts have been constrained by the usually very limited information available in the sampling frames. For this reason there are more studies about attrition in panel surveys, because in a panel one can use information given by the respondents in a previous wave of data collection to explain response behavior in a more recent wave, see for instance, Brose and Klevmarken (1993), Lepkowski and Couper (2002) and Nicoletti and Peracchi (2005). The results from these studies are interesting and important, but they do not necessarily carry over to a cross-sectional survey or the first wave of a panel survey.

It is well-known that response is usually much lower in the first wave of a panel survey than in successive waves and that attrition thus takes place in an already selected sample. People who are notoriously difficult to trace and convince have already been eliminated from the sample in the first wave,¹ see Laurie et al. (1999). Lepkowski and Couper (2002) argue that the response process in the first wave is fundamentally different from that of subsequent waves. This is both because of self-selection of the sample units and because of the extra information and organizational experience gained by the survey agencies at each successive wave. Fitzgerald et al. (1998) reported the same experience from the Panel Study of Income Dynamics (PSID). The attrition between the first and the second wave was 12 percent, for the 20 next waves attrition was on average between 2.5 and 3.0 percent.

Another problem with using data from a previous survey wave to explain response behavior is that survey data always have measurement errors and other types of nonsampling errors. Depending on the variables used this might become a problem when estimating a response model.

In this study we have the advantage of having exceptionally good sample frame data that can be used to explain response behavior. The sample frame was the 2001 wave of the longitudinal register based data set LINDA of Statistics Sweden. LINDA is a random sample including a few hundred thousand individuals from the Swedish population. Register data include population censuses, schooling, income, wealth and tax data, etc. From this source we selected by simple random sampling a smaller sample of 1,430 individuals 50-84 years old to which CATI interviews were administered by Statistics

¹ Depending on design one might try to recruit those who did not participate in the first wave to participate in a second wave, but in many surveys this is never attempted.

Sweden.² These telephone interviews included sequences of questions taken from the U.S. HRS survey and the European SHARE survey and adapted to Swedish circumstances. There were thus questions about health, labor force participation, wages, incomes and wealth. Most of these questions were about “facts” not about feelings, perceptions and attitudes. The average interviewing time was less than 30 minutes. The field work was done in the period April 3 – May 11 2003 with nonresponse follow up June 2-22. In this period most Swedes completed their self assessment for income taxation, so the information needed to answer questions about incomes, assets and taxes should have been timely.

Prior to the field work the questionnaire was tested in the questionnaire laboratory of Statistics Sweden and in a small pretest. Interviewers were experienced telephone interviewer. They got a four hours long training session focusing specifically on our survey and they were afterwards asked to train on the questionnaire before they were allowed to work in the field. The nonresponse follow up was done by a few of the most experienced interviewers.

The contribution of this paper is thus an analysis of the response behavior in a cross-sectional survey with standard questions about health, incomes, taxes and assets using register data that in most cases are considered highly accurate.

² For this age group Linda included 137,557 individuals and the population size was 3,026,499

2. Reasons for Nonresponse

2.1 A literature review

In analyzing the response process in panel surveys Lepkowski and Couper (2002) distinguished between three different conditional processes: Location, contact and cooperation of the sample units. In our case, as in most cases, it might be difficult to make a distinction between location and contact.

Groves and Couper (1998) tried intuitively to explain what determines unit nonresponse. They did that by separately analysing contact and cooperation. Contactability is primarily a function of physical barriers to accessing the respondents, the households' at home pattern, and the interviewers contact schedule. Once a unit is contacted we want them to cooperate. The explanation of cooperation is more complex, it involves the interaction of demographic, socioeconomic and psychological influences.

Groves et al. (1992) presented a theory to explain the survey participation decision. This theory used both observed variables such as socio-demographic properties of the respondent, survey design features and less frequently observed circumstances, such as the interaction between the interviewer and the interviewees. The design properties are known, data about the respondents might be obtainable from the sampling frame, while it is more difficult to get detailed information about the interaction between interviewer and respondent. In some surveys the interviewers are asked to summarize their experiences from each interview, but this is not common practice.

Previous empirical research has suggested that attrition from a panel is more likely for individuals who are on welfare, unmarried, older and nonwhite. Also, attritors have less education, work fewer hours, have lower labor income, and are more likely to rent their homes than the average respondent (Fitzgerald et al. (1998)). Zabel

(1998) concluded that attritors were more likely to live in urban areas, be nonwhite and unmarried, have fewer children and rent their homes. Campanelli et al. (1997) analyzed attrition both on a household level and on an individual level – their main results are in line with the ones above, i.e., respondents who are economically less well off are less likely to be included in the survey. In their literature review Särndal and Lundström (2005) concluded that the response rate is usually expected to be lower among metropolitan residents, single persons, members of childless households, older persons, divorced or widowed persons, persons with low educational attainment, and self-employed persons.

In decomposing attrition into noncontact and refusal Campanelli et al. (1997) found, in line with previous research, that these two groups have different socio-economic characteristics. Nonwhites were harder to contact than whites, as was unmarried respondents compared to married. It was harder to establish contact with young respondents than with old, but once contacted they were generally cooperative. For elderly it is the other way around. Households with no children were more likely to refuse, as were households with many working members, and households consisting of couples.

Lynn et al. (2002) also made a distinct difference between the difficulty of contacting sample members (“ease of contact”) and difficulty of obtaining cooperation once contact is made (“reluctance of cooperation”). In a descriptive analysis based on various health and socio-economic surveys from the UK they found that the probability of participation was not dependent on the number of calls until contact. They also tested the hypothesis that households that were hard to contact have other characteristics than households who were easy to contact. Their main results were that respondents who were hard to contact were more likely to be smokers and drinkers, to have lower blood pressure, be less likely to have a severe illness, be younger, more

likely to be employed and less likely to be white. But there were no significant differences in the characteristics of respondents who were easy to contact and of those who were reluctant respondents.

Finally, Nicoletti and Peracchi (2005) modelled the response behaviour using a bivariate probit model that distinguished between contact and cooperation. They used data from the European Community Household Panel (ECHP). Most of their results are in line with what is expected from previous research. They found that the number of children and home ownership increased the probability of contact, while the number of adults in the household and the equivalised household income (household income divided by the number of household members) both were insignificantly different from zero. They also found that being out of the labor market increased the probability of cooperation whereas being single decreased the same probability. There was no significant effect of the age or education of the respondent.

2.2 Our survey

In the remainder of this section we will discuss contact and cooperation difficulties arising in our survey. Statistics Sweden had mailing addresses to everyone – the address on which the respondents had registered with the tax authorities – and through computerized telephone directories they could get telephone numbers to most of the respondents. However, it is possible to be registered on one address and live somewhere else, for instance old people might have kept their old home while they in fact stay for a longer or shorter period in a nursing home. In this case they might not even have a private telephone. Many Swedes have secondary homes and when they are retired they sometimes live there for longer or shorter periods, not only in the summer. Cell phones have become very common and should in principles increase the chances to reach people, but the telephone directories have not always had full coverage of all cell phone numbers. Some people opt in favor of only having a cell phone and no conventional phone, but this is not as common among elderly people. In our survey contacting people meant to get the right telephone number and then get them on the phone. As usual many attempts were made at varying times of the day and at different days of the week.

After a contact has been established it is very much dependent on the interviewer if it is successful or not. Unfortunately our survey data do not have any information about the interviewers, so it is impossible to estimate any interviewer effects on response. All interviewing was done from the Örebro office of Statistics Sweden and interviewers thus called to all areas of the country. The area in which the respondent lives is thus not confounded with interviewer. Because the CATI system allocated respondents to the interviewers without knowing “the track record of the interviewer” it is a plausible hypothesis that any interviewer effects are independent of effects depending on the characteristics of the respondents. There is though one exception: The

more difficult cases, which remained after the main field period had ended, were in the nonresponse follow up turned over to the most skilled interviewers.

What is possible to do in this study is to model response as a function of the characteristics of the respondents. In explaining the probability for a contact we need variables that capture that some people are more mobile than others and that old people due to old age and sickness, for instance dementia, are difficult to contact. The decision about participating in an interview once contacted depends on the time cost of the respondent and the presence of any competing activities. It also depends on the respondent's understanding for and interest in the issues brought up in the interview and the general purpose of the survey. There is also the concern about invasion of privacy. Even if people are interested in contributing to a health survey many respondents are reluctant to reveal information about wages, incomes and in particular wealth.

3. Descriptive analysis of response frequencies

In the end of May 2003 the response rate was 56.5 percent and the share of refusals 19.6 percent. After the conversion attempts in June total response rate increased to 61.6 percent and the share of not found was reduced by 2.9 percentage units and the share of refusals by 3.0 percentage units. In the end 22.6 percent of the sample members refused and 15.8 percent could not be found. The latter figure, however, includes 12 individuals who were classified as over coverage and should have been eliminated. If this is done the response rate increases to 62.1 percent.³ As shown in Table 1 there is virtually no difference in the response behavior of males and females.

³ These individuals had either died or moved abroad between the day of selection and the day of the interview.

Table 1. Response rates by gender.

Status	Male		Female		Total	
	N	%	N	%	N	%
Responded	412	61.2	469	62.0	881	61.6
Refusals	149	22.2	174	23.0	323	22.6
Not reached	112	16.6	114	15.0	226	15.8
Sample size	673	100.0	757	100.0	1430	100.0

Tables 2-6 exhibit response rates by a number of potentially interesting explanatory variables. First Table 2 shows that response rates are smallest among the youngest (50-55) and the oldest. But this result hides reversed age trends among refusals and not reached. Refusals increase with age while not found seems to be a bigger problem among people below the age of 70.

Table 2. Distribution of response status by age.

Status\ Age group	50-55	56-60	61-65	66-70	71-75	76-80	>80	All
Responded	211 <i>59.60</i>	198 <i>63.46</i>	133 <i>65.84</i>	108 <i>61.71</i>	97 <i>62.18</i>	90 <i>59.60</i>	44 <i>55.00</i>	881 <i>61.6</i>
Refusal	77 <i>21.75</i>	48 <i>15.38</i>	46 <i>22.77</i>	39 <i>22.29</i>	42 <i>26.92</i>	45 <i>29.80</i>	26 <i>32.50</i>	323 <i>22.6</i>
Not reached	66 <i>18.64</i>	66 <i>21.15</i>	23 <i>11.39</i>	28 <i>16.00</i>	17 <i>10.90</i>	16 <i>10.60</i>	10.0 <i>12.50</i>	226 <i>15.8</i>
All	354	312	202	175	156	151	80	1430

Note: Column percent in italics.

Table 3 shows that the response rate increases with increasing schooling. Respondents with only compulsory schooling are both harder to find and to convince.

Table 3. Distribution of response status by education.

Status	Compulsory schooling	High school and at most 2 years of university	More than 2 years of university	Missing value	All
Responded	230 <i>57.21</i>	333 <i>65.04</i>	168 <i>66.93</i>	150 <i>56.60</i>	881 <i>61.6</i>
Refusal	98 <i>24.38</i>	95 <i>18.55</i>	47 <i>18.73</i>	83 <i>31.32</i>	323 <i>22.6</i>
Not reached	74 <i>18.41</i>	84 <i>16.41</i>	36 <i>14.34</i>	32 <i>12.08</i>	226 <i>15.8</i>
All	402	512	251	265	1430

Note 1: Column percent in italics. Note 2: Data on schooling are missing for respondents older than 75 years

Unmarried persons are less likely to respond than married. Most of this difference comes from a higher frequency of not reached for unmarried, while the difference in refusal rate is small.⁴ There is a very clear relation with income, see Table 5. Response rates increase with increasing income. It is most difficult both to contact and to recruit respondents with low incomes.

Table 4. Distribution of response status by marital status.

Status	Married	Unmarried	All
Responded	561 <i>66.63</i>	320 <i>54.42</i>	881 <i>61.6</i>
Refusal	194 <i>23.04</i>	129 <i>21.94</i>	323 <i>22.6</i>
Not reached	87 <i>10.33</i>	139 <i>23.64</i>	226 <i>15.8</i>
All	842	588	1430

Note: Column percent in italics.

⁴ The group unmarried includes people that are cohabiting but not legally married. A similar table but classified by “singles” and “couples”, where the group couples includes married and cohabiting with common children, gave virtually the same result. In the age group 50+ most couples are married.

Table 5. Distribution of response status by monthly income (SEK).

Status	I≤10,000	10,000<I≤15,000	15,000<I≤20,000	I>20,000	All
Responded	156 <i>50.32</i>	136 <i>55.28</i>	242 <i>65.41</i>	347 <i>68.85</i>	881 <i>61.6</i>
Refusal	88 <i>28.39</i>	63 <i>25.61</i>	75 <i>20.27</i>	97 <i>19.25</i>	323 <i>22.6</i>
Not reached	66 <i>21.29</i>	47 <i>19.11</i>	53 <i>14.32</i>	60 <i>11.90</i>	226 <i>15.8</i>
All	310	246	370	504	1430

Note 1: Column percent in italics. Note 2: Income includes labor income and pensions. Pensions have been adjusted by the average compensation rate 0.7 to become comparable to labor incomes

Although there are few respondents that have received any welfare, Table 6 finally suggests that those who are on welfare are difficult to find and also difficult to recruit for an interview.

Table 6. Distribution of response status if on welfare or not.

Status	No welfare	Welfare	All
Responded	875 <i>62.54</i>	6 <i>19.35</i>	881 <i>61.6</i>
Refusal	312 <i>22.30</i>	11 <i>35.48</i>	323 <i>22.6</i>
Not reached	212 <i>15.15</i>	14 <i>45.16</i>	226 <i>15.8</i>
All	842	31	1430

Note: Column percent in italics.

Just by looking at univariate distributions it is difficult to assess which variables are the most important to explain response, because age, schooling, income and welfare are all confounded. We get, however, a very clear message from these tables, namely that response rates are much lower among low skilled and low income people, many of whom are found among the oldest in the sample.

4. A sequential bivariate probit model with univariate selection

The sequence of events we wish to model is first the contact and if contact is established the event of giving an interview. Following Nicoletti and Peracchi (2005) we will use a bivariate probit model. Let Y_1 be a dummy variable that takes the value one if a contact is established and Y_2 another dummy variable that takes the value 1 if an interview is obtained. Assume the following model

$$\begin{aligned} Y_1^* &= \beta_1' X_1 + \varepsilon_1; \\ Y_2^* &= \beta_2' X_2 + \varepsilon_2; \\ Y_1 &= 1 \text{ if } Y_1^* > 0; \text{ otherwise } Y_1 = 0; \\ Y_2 &= 1 \text{ if } Y_1^* > 0 \text{ and } Y_2^* > 0; \text{ otherwise } Y_2 = 0; \end{aligned}$$

where Y_1^* and Y_2^* are bivariate normal latent variables, while ε_1 and ε_2 are bivariate standard normal. The X -vectors are vectors of exogenous explanatory variables uncorrelated with the ε 's.

Conditions are necessary for the above censored bivariate probit model to be identifiable. If the covariates in the contact and the participation equations are the same, then the model is not identified. Identification becomes possible if X_1 and X_2 are not identical, i.e. exclusion restrictions are needed. In this respect we were guided by previous results and common sense. For instance, the variables “if having a leisure home” and “if having stayed in a hospital” were assumed to determine the probability of contact rather than the probability of cooperation. In the final specification a few insignificant variables were deleted from either equation. The model was estimated by maximum likelihood.

Variables assumed to capture the probability of a contact are: The age and schooling of the respondent, the number of household members, marital status, if an immigrant, if received any sickness benefits in 2002, if stayed at a hospital in 2002, if unemployed, if owning a secondary home, if living in a big city, smaller city or in an urban area, the wage rate if working, and if on welfare. The wage rate variable is a monthly rate. It can of course only be observed for those who have a job. It is not obvious how one should define a measure of time cost for those who do not work in the market. We have chosen to introduce a dummy variable that takes the value one if the respondent has no wage income. The wage rate variable takes the value zero for these respondents. An alternative approach is to try to estimate a wage rate for those who do not work had they worked. This can be achieved if a labor supply and an earnings function are estimated jointly with the contact (response) function; for an application to panel data see Brose and Klevmarken (1993).

The decision to give an interview is explained by a subset of the same variables: The age and schooling of the respondent, the number of household members, marital status, if an immigrant, the wage rate if working, and if on welfare.

Descriptive statistics can be found in Table 7. Because the descriptive statistics suggested that the relation with age was not exactly the same for the contacts as for the response once contacted, two different age classifications were used, one in the contact equation and one in the response equation. One might also note that the frequencies for the three educational groups do not sum to one because information about education is missing for respondents older than 75.

Table 7. Descriptive statistics of independent variables.

Variable	Mean	S.D.
Age1 (≤ 55)	0.247	0.432
Age2 (56-75)	0.591	0.491
Age3 (76+)	0.162	0.368
Age4 (≤ 60)	0.466	0.499
Age5 (61-70)	0.263	0.440
Age6 (71+)	0.271	0.444
If female	0.530	0.499
Compulsory	0.281	0.449
High school	0.358	0.479
University	0.175	0.380
Wage	9,490	12,870
No wage	0.462	0.498
Sickness benefit	0.111	0.314
Social security	0.022	0.146
Major city	0.330	0.470
Urban	0.534	0.498
Rural	0.136	0.343
Leisure home	0.141	0.348
Household size	1.894	0.895
Unemployed	0.046	0.210
Married	0.588	0.492
Immigrant	0.042	0.202
Hospital stay	0.111	0.314

The maximum likelihood estimates are presented in Table 8. These results show that the probability of contact increases with age. There is no significant difference between males and females, while couples are easier to contact than singles. People with high school or university are somewhat more difficult to contact than people with only compulsory schooling, but these estimates are uncertain. Those who have sickness benefits have a higher probability of contact (they are at home) while those who have stayed in hospital have a lower probability. People on welfare and immigrants are much more difficult to contact than the average person, and the probability of contact is smaller in the big cities than in other urban and rural areas. People who have a secondary home are also more difficult to contact. Those who do not work have a smaller probability to be contacted than average, which perhaps is counter intuitive. There is no significant effect of being unemployed in addition to not working. The estimate for

the wage rate variable was small and insignificant and thus dropped from the equation.

Table 8. ML estimates of a bivariate probit model.

Variable	<u>Participation given contact</u>		
	Estimate	S.D.	<i>p</i> -value
Constant	0.660	0.270	(0.014)
Age2 (55-75)	0.226	0.104	(0.030)
Age3 (76-)	0.083	0.180	(0.664)
If female	0.018	0.077	(0.819)
High school ^A	-0.197	0.117	(0.093)
University ^B	0.022	0.110	(0.840)
Wage rate	-6.00e-06	4.24e-06	(0.157)
No wage	-0.517	0.112	(0.000)
If welfare benefits	-0.933	0.300	(0.002)
Household size	-0.068	0.063	(0.278)
Married	0.218	0.128	(0.089)
Immigrant	-0.492	0.205	(0.016)
<u>Contact</u>			
Constant	0.473	0.177	(0.008)
Age5 (61-70)	0.496	0.122	(0.000)
Age6 (71-)	0.731	0.162	(0.000)
If female	0.066	0.085	(0.436)
High school ^A	-0.176	0.118	(0.137)
University ^B	-0.031	0.113	(0.782)
No wage	-0.354	0.126	(0.005)
If sickness benefits	0.155	0.145	(0.287)
If welfare benefits	-0.434	0.240	(0.070)
Urban	0.191	0.091	(0.035)
Rural	0.339	0.140	(0.015)
If leisure home	-0.207	0.121	(0.087)
Household size	0.071	0.067	(0.285)
Unemployed	0.100	0.195	(0.606)
Married	0.444	0.117	(0.000)
Immigrant	-0.492	0.180	(0.006)
Hospital stay	-0.280	0.124	(0.008)
Residual correlation	0.597	0.333	

Note: The reference category for education is individuals with compulsory education or older than 75. ^A Includes individuals with a high school degree or individuals who studied at the university less than two years. ^B Includes individuals with more than two years at university.

The probability of a successful contact is higher among people between 55 and 75 years of age than among both younger and older respondents. We do not find any gender effect in this case either. Couples are more cooperative than singles while immigrants and those

on welfare have a much smaller probability than average. Those who do not work have a much smaller probability to cooperate than average and the wage rate effect for those who work is negative suggesting an increasing time cost with increasing wage rate.

Table 9. ML estimates of a bivariate probit model using an alternative definition of the wage rate variable.

Variable	Participation given contact		
	Estimate	S.D.	p-value
Constant	0.612	0.273	(0.025)
Age2 (55-75)	0.205	0.103	(0.047)
Age3 (76-)	0.017	0.179	(0.923)
If female	0.005	0.077	(0.944)
High school ^A	-0.213	0.116	(0.066)
University ^B	0.034	0.109	(0.750)
Wage rate	-4.29e-06	4.92e-06	(0.383)
No wage	-0.366	0.133	(0.006)
Social security	-0.964	0.302	(0.001)
Household size	-0.068	0.063	(0.280)
Married	0.229	0.125	(0.066)
Immigrant	-0.544	0.201	(0.007)
<hr/>			
		Contact	
Constant	0.492	0.178	(0.006)
Age5 (61-70)	0.523	0.129	(0.000)
Age6 (71-)	0.758	0.165	(0.000)
If female	0.049	0.085	(0.561)
High school ^A	-0.177	0.117	(0.129)
University ^B	-0.020	0.112	(0.855)
No wage	-0.357	0.128	(0.005)
Sickness benefit	0.153	0.141	(0.279)
If welfare benefits	-0.442	0.241	(0.066)
Urban	0.187	0.090	(0.037)
Rural	0.339	0.139	(0.015)
Leisure home	-0.206	0.121	(0.089)
Household size	0.071	0.067	(0.286)
Unemployed	0.123	0.198	(0.534)
Married	0.445	0.117	(0.000)
Immigrant	-0.513	0.180	(0.004)
Hospital stay	-0.276	0.124	(0.026)
<hr/>			
Residual correlation	0.628	0.299	

Note 1: Lower threshold for the wage rate variable is 3,000 SEK. Note 2: The reference category for education is individuals with compulsory education or older than 75. ^A Includes individuals with a high school degree or individuals who studied at the university less than two years. ^B Includes individuals with more than two years at university.

The monthly wage variable includes some very small wage rate observations suggesting that these respondents only worked part-time or part of the year. To test the sensitivity of our results for an alternative cut off for this variable the model was re-estimated with the lower threshold of 3,000 SEK instead of zero. The corresponding results are presented in Table 9. The estimate of the wage rate variable is still negative but it moves closer to zero and its standard error increases.

Our model allows for a correlation between the contact and participation equation whereas Lepkowski and Couper (2002) assumed independence. By doing that they assume that omitted variables have no joint impact on contact and participation. A more reasonable assumption is to assume that unobservables influence both the probability of contact and that of participation and thus create a correlation between the contact and participation equations. Results are not conclusive. Using the estimates of Table 8 the correlation is not significantly different from zero while the corresponding test from Table 9 is significant at a level of 5 percent. In both cases the correlation is positive and rather high, approximately 0.6.

4.1. Using register information about diagnosis

In this section we will present results from a reestimated model using more informative diagnosis data. Above we only used a dummy variable indicating if the respondent was admitted to a hospital in 2002, disregarding the reason for admission. Since we have register information about each admitted respondent's diagnosis we can identify those patients who have mental health problems or any type of behavioural disorder. The hypothesis is that those individuals are harder to contact than individuals without any mental problems. We call this variable *if psychiatric diagnosis*.

We know from Table 7 that 11 percent of the respondents were admitted to a hospital for at least one night in 2002, and of those 18 individuals or 1.3 percent of the sample had a psychiatric diagnosis. Although the number of people with this problem is small Table 10 suggests that individuals with a psychiatric diagnosis are difficult to locate and get to participate.

Table 10. Distribution of response status by psychiatric diagnosis.

Status	No psychiatric diagnosis	Psychiatric diagnosis	All
Responded	875 <i>61.97</i>	6 <i>33.33</i>	881 <i>61.6</i>
Refusal	319 <i>22.59</i>	4 <i>22.22</i>	323 <i>22.6</i>
Not reached	218 <i>15.44</i>	8 <i>44.44</i>	226 <i>15.8</i>
All	1412	18	1430

Note: Column percent in italics.

Our data also include historical information about past diagnosis for the period 1984-2002. Using all this historical information would increase the share with a psychiatric diagnosis from 1.3 percent to approximately 8 percent. However it is not obvious that all years contribute useful information. Some of those who got a diagnosis in, for instance, 1984 might have recovered by 2002. For this reason we have only used data for a shorter period, 1997-2002.

A total of 3.4 percent of our sample frame had a psychiatric diagnosis at least once in this period. Many of these individuals had a psychiatric diagnosis more than one year, and some of them also had non-psychiatric diagnosis.

In Table 11 below we see the association between response and having at least one psychiatric diagnosis in 1997-2002. The response rate was below 40 percent for this group. 41 percent could not be reached and 20 percent refused to participate.

Table 11. Distribution of response status by psychiatric diagnosis during 1997-2002.

Status	No psychiatric diagnosis	Psychiatric diagnosis	All
Responded	862 <i>62.42</i>	19 <i>38.78</i>	881 <i>61.6</i>
Refusal	313 <i>22.66</i>	10 <i>20.41</i>	323 <i>22.6</i>
Not reached	206 <i>14.92</i>	20 <i>40.82</i>	226 <i>15.8</i>
All	1381	49	1430

Note: Column percent in italics.

Comparing tables 10 and 11 indicate that the response rate is a little higher when historical data are used. A more distant diagnosis is not as strong an indicator of bad health as a more recent diagnosis. Using all years of data would bring this out even more clearly.

Our model was reestimated using both diagnosis information for 1997-2002 and just for 2002. In the latter case the estimate of the effect of being diagnosed with a psychiatric disease on the probability of contact is very uncertain because of the small number of cases. The point estimate does not change when more years are used but the standard error drops. The results for the whole model are displayed in Table 12.

Table 12. ML estimates of a bivariate probit model with psychiatric diagnosis for the period 1997-2002.

Variable	<u>Participation given contact</u>		
	Estimate	S.D.	p-value
Constant	0.726	0.292	(0.013)
Age2 (55-75)	0.221	0.109	(0.042)
Age3 (76-)	0.061	0.188	(0.745)
If female	0.013	0.079	(0.869)
High school ^A	-0.195	0.120	(0.104)
University ^B	0.030	0.113	(0.793)
Wage rate	-5.77e-06	4.42e-06	(0.191)
No wage	-0.518	0.115	(0.000)
If welfare benefits	-0.923	0.307	(0.003)
Household size	-0.073	0.065	(0.255)
Married	0.196	0.137	(0.150)
Immigrant	-0.473	0.216	(0.028)
	<u>Contact</u>		
	Estimate	S.D.	p-value
Constant	0.480	0.175	(0.006)
Age5 (61-70)	0.474	0.121	(0.000)
Age6 (71-)	0.669	0.159	(0.000)
If female	0.070	0.085	(0.410)
High school ^A	-0.158	0.118	(0.183)
University ^B	-0.032	0.112	(0.755)
No wage	-0.352	0.125	(0.005)
If sickness benefits	0.118	0.148	(0.428)
If welfare benefits	-0.406	0.239	(0.090)
Urban	0.192	0.090	(0.034)
Rural	0.329	0.139	(0.018)
If leisure home	-0.192	0.125	(0.125)
Household size	0.066	0.067	(0.321)
Unemployed	0.126	0.200	(0.529)
Married	0.439	0.117	(0.000)
Immigrant	-0.501	0.179	(0.005)
If psychiatric diagnosis	-0.366	0.201	(0.069)
Residual correlation	0.483	0.408	

Note: The reference category for education is individuals with compulsory education or older than 75. ^A Includes individuals with a high school degree or individuals who studied at the university less than two years. ^B Includes individuals with more than two years at university.

In interpreting the results we find that the probability of contact increases with increasing age of the respondent. There is no significant gender difference. The estimates of the effects of schooling on participation are uncertain but suggest that the probability decreases with increasing schooling. Not working in the market (other than

unemployed) has a strong negative effect which is further strengthened if the family is on welfare. Immigrants and respondents with a psychiatric diagnosis are also more difficult to contact than non-immigrants and normally healthy people. We also note that those who have a secondary home have a lower contact probability than otherwise comparable respondents. Married and cohabiting couples and unemployed have a higher contact probability than singles and employed respectively. There are also regional differences. The contact probability is higher in rural areas and in small cities than in large metropolitan areas.

The probability of participation once contacted is higher in the age bracket 55-75 than among both younger and older respondents. Respondents who do not work in the market, are on welfare and immigrants do not only have low contact probabilities but also low participation probabilities. The wage rate captures time cost. Although the estimate is rather uncertain it suggests that the higher wage rate the lower willingness to participate. Schooling is less important in determining participation than contact but there is an indication of a negative effect of just having graduated from high school. Finally, we also note that married and cohabiting are more willing to participate than singles, and that the residual correlation is positive but there is a high standard error. A positive residual correlation implies that there are unobserved factors that influence the two probabilities in the same direction.

5. Concluding remarks

From an economists perspective it might be reasonable to believe that time cost has a strong influence on the probability of contact and participation, and consequently that high wage earners and high income people are difficult to convince to participate in surveys. Confirming previous results about the attrition in panel studies this

study shows that this notion is largely false. It is true that we have found a time cost effect on participation but the major finding is that nonresponse primarily comes from the left tail of the income distribution. Respondents without work, on welfare, immigrants and singles, and respondents living in the major metropolitan areas are those who both are difficult to contact and to convince to participate.

This result would seem to have implications both for survey design and post survey compensation measures. The characteristics of the respondents that contribute to nonresponse suggest that this is a group which is rather uninterested in the research purpose of our survey and that measures should be taken to try to wake up a greater interest. The properties of those who do not respond also suggest that this is a group in an economic situation such that they should be sensitive to economic incentives even if they are rather small.

In addition to the major group of nonresponding, contact efforts should also focus on people who have more than basic training, are in the peak of their work career and have a secondary home.

Our study also suggests that calibration methods that try to compensate for nonresponse should use variables and population information that capture these groups. Gender is, for instance, not such a variable while labor force participation, if immigrant and health status are such variables.

References

- Brose, P. and Klevmarken, A.: 1993, Modeling response in panel survey, *Working Papers of the European Scientific Network on Household Panel Studies*. Paper 81. Colchester: University of Essex. (Presented at the 8th session of the International Statistical Institute, Cairo).
- Campanelli, P., Purdon, S. and Sturgis, P.: 1997, *Can you hear me knocking: an investigation into the impact of interviewers on survey response rates*, Social and community planning research, London.
- Fitzgerald, J., Gottschalk, P. and Moffit, R.: 1998, An analysis of sample attrition in panel data: the Michigan panel study of income dynamics, *Journal of human resources* **33**(2), 251-299.
- Groves, R.M., Cialdini, R.B. and Couper, M.P.: 1992, Understanding the decision to participate in a survey, *Public Opinion Quarterly* **56**(4), 475-495.
- Groves, R.M. and Couper, M.P.: 1998, *Nonresponse in household interview surveys*, John Wiley & Sons, Inc. New York.
- Laurie, H., Smith, R. and Scott, L.: 1999, Strategies for reducing nonresponse in a longitudinal panel survey, *Journal of official statistics* **15**(2), 269-282.
- Lepkowski, J.M. and Couper, M.P.: 2002, Nonresponse in the Second Wave of Longitudinal Household Surveys, in Groves, R.M., Dillman, D.A., Eltinge J.L. and Little R.J.A., *Survey Nonresponse*, John Wiley & Sons, Inc. New York.
- Lynn, P., Clarke, J., Martin, J. and Sturgis, P.: 2002, The effects of extended interviewer efforts on nonresponse bias, in Groves, R.M., Dillman, D.A., Eltinge J.L. and Little R.J.A., *Survey Nonresponse*, John Wiley & Sons, Inc. New York.
- 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**(4), 763-781.
- Lundström, S. and Särndal, C-E.: 2005, *Estimation in surveys with nonresponse*, John Wiley & Sons, Ltd. West Sussex, England.

Zabel, J.E.: 1998, An analysis of attrition in the panel study of income dynamics and the survey of income and program participation with an application to a model of labor market behaviour, *Journal of human resources* **33**(2), 480-506.

WORKING PAPERS*

Editor: Nils Gottfries

- 2004:16 Iida Häkkinen, Do University Entrance Exams Predict Academic Achievement? 38 pp.
- 2004:17 Mikael Carlsson, Investment and Uncertainty: A Theory-Based Empirical Approach. 27 pp.
- 2004:18 N. Anders Klevmarken, Towards an Applicable True Cost-of-Living Index that Incorporates Housing. 8 pp.
- 2004:19 Matz Dahlberg and Karin Edmark, Is there a “Race-to-the-Bottom” in the Setting of Welfare Benefit Levels? Evidence from a Policy Intervention. 34 pp.
- 2004:20 Pär Holmberg, Unique Supply Function Equilibrium with Capacity Constraints. 31 pp.
- 2005:1 Mikael Bengtsson, Niclas Berggren and Henrik Jordahl, Trust and Growth in the 1990s – A Robustness Analysis. 30 pp.
- 2005:2 Niclas Berggren and Henrik Jordahl, Free to Trust? Economic Freedom and Social Capital. 31 pp.
- 2005:3 Matz Dahlberg and Eva Mörk, Public Employment and the Double Role of Bureaucrats. 26 pp.
- 2005:4 Matz Dahlberg and Douglas Lundin, Antidepressants and the Suicide Rate: Is There Really a Connection? 31 pp.
- 2005:5 Maria Vredin Johansson, Tobias Heldt and Per Johansson, Latent Variables in a Travel Mode Choice Model: Attitudinal and Behavioural Indicator Variables. 31 pp.
- 2005:6 Katarina Nordblom and Henry Ohlsson, Tax Avoidance and Intra-Family Transfers. 25 pp.
- 2005:7 Sören Blomquist and Luca Micheletto, Optimal Redistributive Taxation when Government’s and Agents’ Preferences Differ. 22 pp.
- 2005:8 Ruth-Aïda Nahum, Income Inequality and Growth: A Panel Study of Swedish Counties 1960-2000. 39 pp.
- 2005:9 Olof Åslund and Peter Fredriksson, Ethnic Enclaves and Welfare Cultures – Quasi-experimental Evidence. 37 pp.
- 2005:10 Annika Alexius and Erik Post, Exchange Rates and Asymmetric Shocks in Small Open Economies. 31 pp.

* A list of papers in this series from earlier years will be sent on request by the department.

- 2005:11 Martin Ågren, Myopic Loss Aversion, the Equity Premium Puzzle, and GARCH. 34 pp.
- 2005:12 Pär Holmberg, Numerical Calculation of an Asymmetric Supply Function Equilibrium with Capacity Constraints. 18 pp.
- 2005:13 Jovan Zamac, Winners and Losers from a Demographic Shock under Different Intergenerational Transfer Schemes. 44 pp.
- 2005:14 Peter Welz and Pär Österholm, Interest Rate Smoothing versus Serially Correlated Errors in Taylor Rules: Testing the Tests. 29 pp.
- 2005:15 Helge Bennismarker, Kenneth Carling and Bertil Holmlund, Do Benefit Hikes Damage Job Finding? Evidence from Swedish Unemployment Insurance Reforms. 37 pp.
- 2005:16 Pär Holmberg, Asymmetric Supply Function Equilibrium with Constant Marginal Costs. 27 pp.
- 2005:17 Pär Holmberg, Comparing Supply Function Equilibria of Pay-as-Bid and Uniform-Price Auctions. 25 pp.
- 2005:18 Anders Forslund, Nils Gottfries and Andreas Westermarck, Real and Nominal Wage Adjustment in Open Economies. 49 pp.
- 2005:19 Lennart Berg and Tommy Berger, The Q Theory and the Swedish Housing Market – An Empirical Test. 16 pp.
- 2005:20 Matz Dahlberg and Magnus Gustavsson, Inequality and Crime: Separating the Effects of Permanent and Transitory Income. 27 pp.
- 2005:21 Jenny Nykvist, Entrepreneurship and Liquidity Constraints: Evidence from Sweden. 29 pp.
- 2005:22 Per Engström och Bertil Holmlund, Jenny Nykvist, Worker Absenteeism in Search Equilibrium. 35pp.
- 2005:23 Peter Hästö och Pär Holmberg, Some inequalities related to the analysis of electricity auctions. 7pp.
- 2006:1 Jie Chen, The Dynamics of Housing Allowance Claims in Sweden: A discrete-time hazard analysis. 37pp.
- 2006:2 Fredrik Johansson and Anders Klevmarken, Explaining the size and nature of response in a survey on health status and economic standard. 25pp.

See also working papers published by the Office of Labour Market Policy Evaluation
<http://www.ifau.se/>