ISER Working Paper Series

An experimental analysis of the impact of survey design on measures and models of subjective wellbeing

Steve Pudney
Institute for Social and Economic Research
University of Essex

No. 2010-20
Jun-2010

INSTITUTE FOR SOCIAL
& ECONOMIC RESEARCH

E·S·R·C
ECONOMIC
& SOCIAL
RESEARCH
COUNCIL

www.iser.essex.ac.uk

Non-technical summary

The panel survey *Understanding Society* is a major new resource for economic and social research. It is an important source of information on people's perceptions of their own wellbeing, and is widely used for policy research. Survey questions on subjective wellbeing are contentious, and some researchers and policy commentators are sceptical about their ability to reveal 'true' levels of wellbeing. A commonly-expressed objection is that the responses to questions of this kind are unduly influenced by apparently minor details of survey design, such as question wording, method of delivering the question, and context of the interview.

One of the innovative features of *Understanding Society* is that part of the sample (known as the *Innovation Panel* or *IP*) is reserved for experimental work to examine methodological issues such as this. We have used the second wave of the *IP*, conducted in 2009, to carry experiments designed to investigate the influence of these design details on responses to questions about various domains of satisfaction: specifically the respondent's satisfaction with his or her state of health; family income; available leisure time; job; and with life overall. Each of the sampled households (which had initially been randomly selected for inclusion in the *IP*) was randomly assigned to one of ten experimental treatment groups, each of which received a different 'treatment', consisting of a different version of question wording, method of asking questions, ordering of questions, or use of visual aids (such as showcards or computer displays).

We find strong evidence that some of these differences in survey design have an influence on the responses that interviewees give to the satisfaction questions and on the results of analyses made using the data from the *IP*.

The three main conclusions are as follows.

(1) Certain aspects of survey design, such as the greater privacy achieved through computer-based self completion questions and some details of question wording and format, have a large and significant effect on the pattern of responses given by survey participants.

(2) The responses made by women appear to be more strongly influenced by these design features than those made by men.

(3) When the data are used in an attempt to uncover the relationship between personal circumstances and satisfaction, the research findings that result are not very robust, in the sense that apparently small differences in design features of the survey can make an important difference to the conclusions. This is particularly so for estimates of the influence of health, income, gender, family size and hours of work on various domains of satisfaction. This happens because people in different circumstances are influenced in different ways by these design features, causing distortion of the relationship that we find in the data. The behavioural basis for these heterogeneous effects is unclear, so this is a challenge for future methodological research.

An experimental analysis of the impact of survey design on measures and models of subjective wellbeing

Stephen Pudney
Institute for Social and Economic Research
University of Essex

This version: June 23, 2010

Abstract

We analyse the results of experiments on aspects of the design of questionnaire and interview mode in the 2009 wave of the new UK *Understanding Society* panel survey. The randomised experiments relate to job- and life-satisfaction questions and vary the labeling of response scales, the mode of interviewing and the location of questions within the interview. We find a highly significant impact of these design features on the distributions of reported satisfaction in various life domains and some important impacts on the findings from conventional cross-section models of satisfaction.

Keywords: Survey design, wellbeing, satisfaction, response bias, Understanding Society

JEL classifications: C23, C25, C81, J28

Contact: Steve Pudney, ISER, University of Essex, Wivenhoe Park, Colchester, CO4 3SQ, UK; tel. +44(0)1206-873789; email spudney@essex.ac.uk

This work was supported by the UK Longitudinal Studies Centre and the ESRC Research Centre on Micro-Social Change. I am grateful to Noah Uhrig and Jon Burton for their help with the IP data and to Mark Bryan, Nick Buck, Peter Lynn and Holly Sutherland for helpful comments and suggestions.

1 Introduction

It is well established in the survey methods literature that the way you ask a question may have a big influence on the answer that you get. Although there has been recent interest in reliability issues for subjective wellbeing data (Kristensen and Westergaard-Nielsen 2007 and Krueger and Schkade 2008), the economics literature on happiness and wellbeing has largely neglected the influence of survey design and context and its possible implications for data analysis. However, Conti and Pudney (2010) recently analysed quasi-experimental evidence arising from variations over time in the design of job satisfaction questions in the British Household Panel Survey (BHPS), finding evidence of a substantial influence of the design of questions and response scales, the mode of interview and the interview context on the distribution of reported satisfaction. One of the most striking aspects of the BHPS evidence is the significantly different impact that survey design features have on the response behaviour of men and women and the consequent large distortions that may be induced in research findings on gender differences in the determinants of job satisfaction.

Our aim in this paper is to explore this issue further using evidence from a randomised experiment in the successor to the BHPS, called *Understanding Society* (USoc) and to expand the scope of the work by considering several domains of life satisfaction in addition to job satisfaction and a wider range of variations in survey design.

2 Experimental design

2.1 Treatment groups and survey mode

USoc is a new, very large-scale, household panel survey, which will absorb the BHPS sample from the 2010-11 wave onwards. The design of USoc differs in many ways from that of BHPS and one of its innovative features is the existence of a 1500-household sub-panel, known as the *Innovation Panel* (IP), reserved exclusively for experimental work. Fieldwork

for the second wave of the IP was conducted in April-June 2009, and included experiments to investigate the impact of question design, interview mode and position of questions within the interview, using random assignment of households to treatment groups. This proceeded in two stages. First, households were assigned (in the ratio 2:1) to groups interviewed by Computer-Assisted Telephone Interviewing (CATI) or alternatively face-to-face (F2F) during an interviewer visit to the home. During F2F interviews, most questions were delivered by Computer-Assisted Personal Interviewing (CAPI), but Computer-Assisted Self-Interviewing (CASI) was used for the satisfaction questions in a randomly-assigned subgroup. There were also independent assignments to treatment groups formed by varying question design and position of the question within the interview. As part of the design, for assignments that would have resulted in a requirement to administer by CATI a question that was in fact infeasible by telephone (because it required a showcard or reading of a long list of allowable responses), the closest feasible approximation to the allocated treatment was substituted. Finally, in some cases telephone contact was unsuccessful, in which case the household was instead interviewed F2F, if that proved possible. The resulting treatment groups A-J (defined on an intention-to-treat basis) are detailed in Table 1. Note that individuals within the same household always received the same experimental treatment. In total, 1,788 individuals in 1,100 households were asked the life satisfaction questions, 841 were interviewed by CAPI or CASI during interviewer visits and 947 by telephone. There were no variations in question position within the interview for job satisfaction, so four combined treatment groups (C+G), (D+H), (E+I) and (F+J) were used alongside groups A and B for the job satisfaction experiments. Full details of question structure and placement are given in the next section.

Table 1 Experimental treatment groups

Group	Interview mode	Response scale	Position of question	Sample n [1]	
				F2F[2]	CATI
Job satisfaction and Life satisfaction questions					
A	CASI	Full labels: 1-stage	Late	167	0
B	CASI	Polar labels	Late	136	0
C	CAPI+showcard	Full labels: 1-stage	Late	55	0
D	CATI[3]	Full labels: 2-stage	Late	78	301
E	CAPI+showcard	Polar labels	Late	69	0
F	CATI[3]	Polar labels	Late	68	323
Life satisfaction questions only					
G	CAPI+showcard	Full labels: 1-stage	Early	65	0
H	CATI[3]	Full labels: 2-stage	Early	67	145
I	CAPI+showcard	Polar labels	Early	56	0
J	CATI[3]	Polar labels	Early	69	167

[1] Nos. of responses to question on overall life satisfaction; [2] F2F = CAPI or CASI; [3] CAPI substituted in some cases

2.2 Design of questions and response scales

Questions were asked sequentially for three aspects of life satisfaction, using (for all groups except D and H) the following format:

> *How dissatisfied or satisfied are you with the following aspects of your situation: (a) your health; (b) the income of your household; (c) the amount of leisure time you have.*

These three domain-specific questions were followed by an overall assessment:

> *Using the same scale, how dissatisfied or satisfied are you with your life overall?*

For groups C, E, G and I, a fully-labeled showcard specified response options in a vertical list ordered from top to bottom as: *7 Completely satisfied; 6 Mostly satisfied; 5 Somewhat satisfied; 4 Neither satisfied nor dissatisfied; 3 Somewhat dissatisfied; 2 Mostly dissatisfied; 1 Completely dissatisfied.* For groups A and B, the questions were administered by the more private Computer-Assisted Self-Interviewing (CASI) method, and the seven alternatives were

displayed horizontally across the screen of a laptop computer for selection directly by the respondent. Polar-point labeled variants of the question (groups B, E and I) omitted the textual labels from options 2 to 6. If the polar-labeled response scale was communicated orally, explanations of the two extreme points were read out by the interviewer.

Treatment groups D and H received a question designed specifically for telephone interviewing, for which a showcard would have been impossible and a single question with full list of responses was judged impractical. The question is fully-labeled, with a two-stage structure:

> (*i*) *How dissatisfied or satisfied are you with your life overall? Would you say you are... (1 Dissatisfied; 2 Neither Dissatisfied nor Satisfied; 3 Satisfied)*
>
> (*ii*) [If dissatisfied or satisfied...] *And are you Somewhat, Mostly or Completely [Satisfied / Dissatisfied] with your life overall? (1 Somewhat; 2 Mostly; 3 Completely)*

In addition to the life satisfaction questions, people in employment or self-employment were also asked about their job satisfaction. The standard one-stage question was:

> *All things considered, which number best describes how dissatisfied or satisfied you are with your job overall?*

The same 1-7 response scale and labeling options were used as for the single-stage life satisfaction questions. For groups D and H, the two-stage formulation was:

> (*i*) *All things considered, would you say that you are dissatisfied, neither dissatisfied nor satisfied, or satisfied with your job overall?*
>
> (*ii*) [If dissatisfied or satisfied...] *And would you say that you are you Somewhat, Mostly or Completely [Satisfied / Dissatisfied] with your present job overall?*

Life satisfaction questions were either asked early (about 25% of the way through the interview, following a block of questions on transport mode choices) or late (about 95% of the way through the interview, following questions on political affiliation and values). The job satisfaction question was always asked shortly after mid-interview, following a section dealing with employment or self-employment details, including occupation, hours and earnings.

3 The impact of survey design on data distributions

We begin by testing the hypothesis that there is no difference in the distribution of the satisfaction variable between treatment groups, or between combinations of treatment groups defined by a particular characteristic of the interview mode or question design, using the nonparametric Kruskal-Wallis (K-W) test, the P-values for which are presented in Table 2. In view of Conti and Pudney's (2009) findings on gender differences, we carry out these tests separately for men and women. To avoid sample dependence, only one male and one female is selected (randomly) from households with multiple male or female members. The overall test of the hypothesis of identical response distributions across all treatment groups (row 1 of Table 2) yields significant (at the 5% level) results for all except the income satisfaction variable among women. For men, the only significant differences are for satisfaction with life overall and job satisfaction. Using a Bonferroni adjustment for multiple testing across the five satisfaction variables, we would find significant evidence of survey design impacts for women but not for men (at an overall 5% significance level). To indicate the possible source of these response differences, we also calculate K-W tests for contrasts focusing on response labeling, interview privacy, question structure, question placement and interview mode, which suggests several conclusions.

First, the early or late positioning of questions within the questionnaire is not associated with any significant shifts in the response distribution. This is in contrast to the large questionnaire context effects that have been found in many other survey applications (Schuman

and Presser 1981, Tourangeau 1999) and the evidence of respondent fatigue which may affect responses late in the interview (Herzog and Bachman 1981, Helgeson and Ursic 1994). Note that we do not investigate the ordering of individual questions within the satisfaction module – something that has been found to influence respondents' interpretation of satisfaction questions (Schwarz et al 1991, Tourangeau et al 1991).

Second, we find that the labeling of points on the response scale has a significant impact on the pattern of responses, in line with much other evidence in the survey methods literature (Weng 2004), and we also confirm Conti and Pudney's (2010) finding for the BHPS job satisfaction questions that this effect is larger for female respondents than for males.

Third, the method of question delivery (CAPI v CASI and CAPI/CASI v CATI) has a significant impact for women in relation to satisfaction with leisure, life overall and job but, for men, the result is significant only in one case (CASI v CAPI for satisfaction with life overall). It is hard to interpret these mode effects, since they involve differences in several dimensions, including the format of visual display of the response scale (Jenkins and Dillman 1997), the degree of respondent privacy and presence of an outsider (the interviewer). Privacy and the social desirability of alternative responses are especially important for sensitive issues (Hochstim 1967, De Leeuw 1992, Aquilino 1997) and a further important factor may be a desire by some individuals to maintain a bargaining position within the family, rendering some satisfaction questions sensitive in oral interviews where other family members may be within earshot (Conti and Pudney 2010).

Fourth, there is a significant difference between the response distributions produced by the one-stage and two-stage question structures, again much more significant for women (all satisfaction variables except income) than for men (job satisfaction only). There has been some debate about the use of two-stage branching (or unfolding) question structures, some authors finding better reliability for them (Krosnick and Berent 1993),[1] while others have

[1]Note that differences in question structure were confounded with labeling differences in the Krosnick-Berent study of test-retest reliability. I would also argue that test-retest reliability should be seen as a

found that some respondents are unable to interpret the question in the intended manner without access to the full range of allowed responses (Hunter 2005, p.10-11).

Overall, it is striking that women's response behaviour appears more sensitive than men's to multiple aspects of the interview process, although the behavioural basis for these differences is not clear. Whatever the source, the implications for the analysis of gender differences are potentially serious.

Table 2 Kruskal-Wallis tests for equality of response distributions (unadjusted P-values)

	Health		**Income**		**Leisure**		**Life overall**		**Job**	
Group	Women	Men	Women	Men	Women	Men	Women	Men	Women	Men
All	0.031	0.187	0.105	0.367	0.030	0.112	0.000	0.011	0.003	0.036
CASI v CAPI	0.119	0.348	0.248	0.387	0.013	0.115	0.001	0.021	0.002	0.519
Labeling	0.015	0.290	0.032	0.795	0.007	0.570	0.002	0.109	0.003	0.173
1- v 2-stage	0.013	0.711	0.137	0.959	0.001	0.270	0.000	0.126	0.000	0.031
Early v late	0.702	0.166	0.650	0.065	0.463	0.559	0.643	0.144	-	-
CAPI/CASI v CATI	0.167	0.174	0.382	0.838	0.042	0.269	0.001	0.906	0.044	0.136

What form do the significant impacts take? Table 3 presents summary statistics for each of the five satisfaction variables, denoted Y_H, Y_I, Y_L, Y_O and Y_J for health, income, leisure, life overall and job respectively: the sample mean and the sample proportions of the events $Y_d \geq 6$ ("mostly or completely satisfied") and $Y_d = 7$ ("completely satisfied") for domain d.[2] Note that many researchers have used a binary indicator of the event $Y_d \geq 6$ as the dependent variable in models of satisfaction. The differences in response distributions across treatment groups turn out to be quite complex. For three of the four contrasts the sign of the difference in the mean is uniform across all five satisfaction variables. Polar-point labeling, CAPI/CASI interviewing and one-stage question design tend to reduce the mean level of reported satisfaction relative to the alternatives of full labeling, telephone interviewing and

measure of consistency over time rather than a true statistical reliability measure such as absence of bias or MSE.

[2]We do not give a gender breakdown of Table 3, but there are few qualitative differences in the conclusions for the male and female samples (details available from the author).

two-stage question wording. The effect is less clear for CASI v CAPI, where the differences go in different directions for health and income than for leisure, life overall and job.

These relatively uniform differences in the mean conceal complex differences in the underlying response distributions. Consider the case of response category labeling: the effect of polar-point labeling is, as expected, to increase substantially the proportion of extreme ($Y_d = 7$) responses, for each of the five satisfaction variables, since the label draws the eye towards the extreme point. However, this is not a simple shift of probability from $Y_d = 6$ to $Y_d = 7$, since there is also a large reduction in the sample proportion of $Y_d \geq 6$ for all satisfaction domains. Figure 1 shows the difference in distributional shape, for the example of overall life satisfaction, and it is evident that polar-point labeling tends to increase the sample proportion of $Y_d = 5$ ("somewhat satisfied") alongside the reduction for $Y_d = 6$ and increase for $Y_d = 7$. This complex of differences does not have any obvious simple interpretation.

Table 3 Summary statistics by treatment group

Summary	**Labeling**		**Privacy**		**Mode**		**Structure**	
statistic	Full	Polar	CASI	CAPI	F2F	Tel	1-stage	2-stage
Health satisfaction								
$\bar{Y}_H$	5.33	5.123	5.22	5.16	5.18	5.35	5.26	5.30
$P(Y_H \geq 6)$	0.653	0.504	0.566	0.547	0.554	0.561	0.540	0.593
$P(Y_H = 7)$	0.161	0.217	0.152	0.235	0.205	0.249	0.204	0.277
Income satisfaction								
$\bar{Y}_I$	4.84	4.70	4.74	4.71	4.72	4.79	4.74	4.80
$P(Y_I \geq 6)$	0.455	0.342	0.414	0.397	0.403	0.386	0.370	0.441
$P(Y_I = 7)$	0.108	0.129	0.099	0.157	0.136	0.138	0.124	0.163
Leisure satisfaction								
$\bar{Y}_L$	5.02	4.94	4.84	4.97	4.92	5.17	4.96	5.24
$P(Y_L \geq 6)$	0.493	0.439	0.439	0.494	0.474	0.517	0.453	0.584
$P(Y_L = 7)$	0.203	0.276	0.193	0.284	0.250	0.299	0.258	0.312
Overall satisfaction								
$\bar{Y}_O$	5.57	5.48	5.30	5.56	5.47	5.67	5.51	5.71
$P(Y_O \geq 6)$	0.685	0.579	0.546	0.647	0.610	0.665	0.605	0.707
$P(Y_O = 7)$	0.164	0.224	0.152	0.252	0.215	0.259	0.209	0.297
Job satisfaction								
$\bar{Y}_J$	5.27	5.20	5.08	5.23	5.17	5.49	5.22	5.57
$P(Y_J \geq 6)$	0.586	0.481	0.520	0.544	0.535	0.580	0.508	0.656
$P(Y_J = 7)$	0.130	0.207	0.162	0.216	0.195	0.267	0.187	0.320

Use of a two-stage rather than one-stage question design also produces uniform differences for each of the three summary measures. It generates a higher mean response and a substantially higher frequency of $Y_d = 7$ responses but, in contrast to polar-point labeling, it also produces the anticipated increase in the $Y_d \geq 6$ frequency. Inspection of the empirical distribution (Figure 1) suggests an interpretation of this as the result of an aversion to the lowest response point at each stage of questioning, causing avoidance of $Y_d = 5$. There is also an increase in the $Y_d = 4$ frequency, corresponding to the "neither dissatisfied nor satisfied" response at stage 1, reflecting the well-known bias in favour of the central category in scales involving a neutral central point.

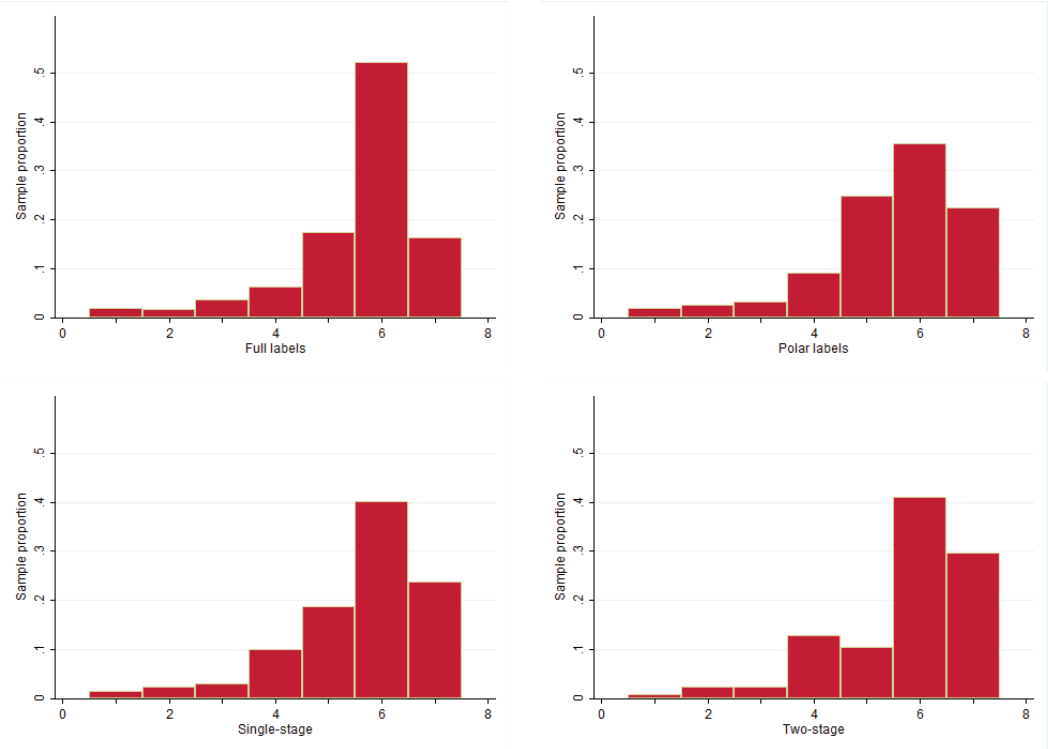

Figure 1 Distributions of responses to overall life satisfaction question: Full v polar labeling; and 1-stage v 2-stage design

The CASI v CAPI and F2F v Telephone contrasts do not generate uniform differences across satisfaction domains, except for the sample proportion of the extreme "completely satisfied" response, which is increased by the use of CAPI rather than CASI and by telephone

rather than F2F interviewing. In all but one case (satisfaction with income), telephone interviewing also increases the proportion of "completely" or "mostly satisfied" responses. The empirical distributions for CASI v CAPI and F2F v telephone for the example of overall life satisfaction are shown in Figure 2.

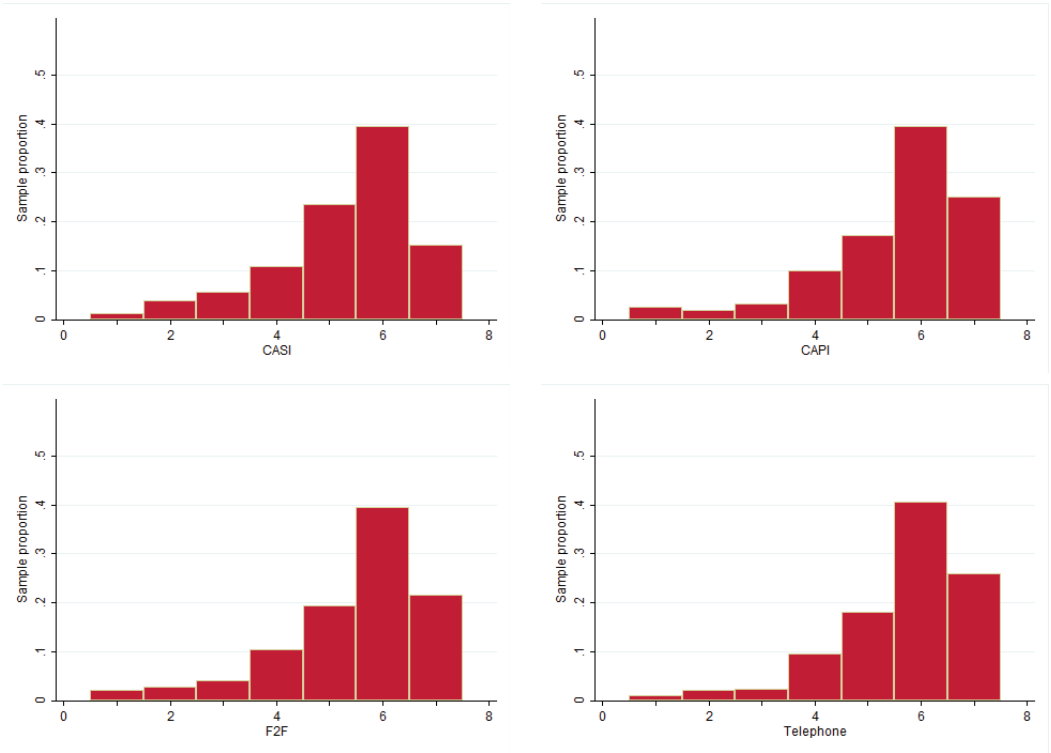

Figure 2 Distributions of responses to overall life satisfaction question: CASI v non-CASI; and CAPI/CASI v telephone

4 The impact of survey design on models of satisfaction

The demand for data is a derived demand – we are interested in data only because of the research results that can be produced from them. Much of the survey methods literature ignores this fundamental point and concentrates on the impact of design features on the statistical reliability of relatively simple summary measures computed from the data. In contrast, most applied researchers are interested in the statistical relationships between variables, using models which represent some conditional distribution in the data. In the research literature on wellbeing, this type of modeling takes the form of relationships between sat-

isfaction as a dependent variable and a large set of covariates describing the individual's characteristics and circumstances (see Van Praag and Ferrer-i-Carbonell 2004, and Clark et al 2008 for surveys). Typical analysis methods include ordered probit, linear regression and binary probit. We apply these modeling approaches and investigate the impact of experimental variations in survey design on the estimates. The covariates used in these models are intended to be broadly representative of those appearing in the published literature and are described in Appendix Table A1.

As a first step, we test for the significance of design features by adding treatment group dummy variables as additional covariates. Appendix Tables A2-A4 give gender-specific results for the five satisfaction variables, using ordered probit, linear regression and binary probit for the event $Y_d \geq 6$. In terms of the influence of personal and household characteristics and circumstances, the findings from this modeling exercise are mostly quite standard. Satisfaction generally increases with health; there is a U-shaped age profile; broken relationships and cohabitation rather than marriage are associated with lower satisfaction; members of non-white minorities report lower satisfaction; unemployment has a large negative impact; the arrival of a new child is associated with lower satisfaction (particularly with income and leisure); and there is a negative satisfaction-education gradient. However, these effects are not always statistically significant, owing to the small sample sizes in some treatment groups.

Two features invite comment: first, for health status, we use a rich set of covariates: a categorical self-assessment of general health (represented as a set of category-specific dummy variables), a binary variable for the existence of a self-reported disability or difficulty with basic activities, and a dummy recording the existence of a past diagnosis of one or more of a set of serious illnesses, including diabetes, heart disease, cancer, etc. For satisfaction with health, all of these variables have significant coefficients, with a surprising positive effect for diagnosed illness – suggesting that a formal diagnosis (and the treatment that goes with it) has some reassuring effect. For other domains of satisfaction, only self-assessed

health status is ever significant and the two other health variables have been dropped from the reported models. Second, income effects are represented by gross household income per head and the respondent's personal income as a share of the household total. For all satisfaction domains except the job, the only significant income effect is a negative impact of the personal income share among women. This could be interpreted either as dissatisfaction with the responsibility of being the family's main source of financial support or aversion to the greater risk of a non-diversified income portfolio. For job satisfaction, the picture is quite different, with both household income and the respondent's income share having a positive effect. This emphasises the possibility of conflicting feelings: that someone may get a sense of 'professional' satisfaction from being a primary earner, whilst being dissatisfied in a wider sense with the responsibility or risk that position brings. We also find confirmation here of the common finding that men's job satisfaction is more sensitive to the wage than is women's, while the reverse is true for working hours (Conti and Pudney 2010).

Table 4 summarises the significance of the fixed additive treatment effects by means of Wald test statistics for their joint significance. Note that the individual tests comprising Table 4 are not statistically independent, and that attempts to draw conclusions spanning satisfaction domains, analysis methods or gender groups are complicated both by this dependence and by the fact that multiple comparisons are involved. Use of adjustments such as the Bonferroni approximation implies the use of smaller significance cutoffs for these P-values than would be the case for individual tests. Bearing this in mind, the evidence of an impact of survey design remains stronger for women than for men. Among the five domains of satisfaction, the impact is clearest for job satisfaction (for both sexes) and satisfaction with leisure and with life overall (for women). A clear finding from the test results is that the method of analysis matters. The common practice of using binary probit for "mostly" or "completely satisfied" ($Y_d \geq 6$) has the effect of focusing on a specific part of the response distribution which, in some cases, appears to be particularly vulnerable to influence from aspects of survey design.

Table 4 Significance of experimental treatment effects (P-values for Wald tests on coefficients of treatment group dummies)

Satisfaction variable	**Ordered probit**		**Linear regression**		**Binary probit**	
	Female	Male	Female	Male	Female	Male
Health	0.0761	0.0611	0.3112	0.0114	0.0015	0.0000
Income	0.3420	0.4476	0.4664	0.4944	0.0090	0.0456
Leisure	0.0011	0.0915	00018	0.1617	0.0000	0.0066
Life overall	0.0016	0.1336	0.0038	0.3357	0.0000	0.0164
Job	0.0087	0.0583	0.0229	0.0971	0.0061	0.0236

The models underlying the tests presented in Table 4 incorporate the assumption that the coefficients of explanatory covariates are design-invariant, which we now investigate. It is not feasible to estimate separate models for each of the treatment groups because of the limited size of each group. For the same reason, it is not possible to investigate the stability of all the coefficients with respect to design features. To avoid proliferation of models, we analyse combined samples of males and females, with gender entering both as an additive dummy and through gender-specific coefficients for the following variables: household income and personal income share, the wage working hours and (for leisure satisfaction only) the number of dependent children. We then test for the existence of differences in particular coefficients of interest across two sets of treatment groups which represent contrasts in some aspect of survey design. This is done separately for four design aspects: response labeling, interview privacy, interview mode and question structure.[3] The model specification is an ordered probit, based on the following latent regression:

$$Y^{\star}_{id} = \boldsymbol{x}_i\boldsymbol{\beta}_d + \boldsymbol{x}^{(d)}_i\zeta_i\boldsymbol{\gamma}_d + \zeta_i\delta_d + u_{id}\ , \qquad d = H, I, L, O, J \tag{1}$$

where $Y^{\star}_{id}$ is the latent variable underlying individual i's reported level of satisfaction with the dth life domain, $\boldsymbol{x}_i$ is the full vector of explanatory variables, $\boldsymbol{x}^{(d)}_i$ is the subset of

[3]Respectively, these four contrasts compare composite groups: (A+C+D+G+H) and (B+E+F+I+J); (A+B) and (C+D+E+F+G+H+I+J); (CAPI+CASI) and CATI; and (A+B+C+E+F+G+I+J) and (D+H). Note that the third of these contrasts is based on a division involving self-selection, since there was some transfer, in the field, of households experimentally assigned to CATI to the CAPI mode, as a result of contact difficulties. All other contrasts are based on experimental assignments. Using instead an 'intention-to-treat' basis for the third contrast (in other words, (A+B+C+E+G+I) v (D+F+H+J)) does not alter the results in any important way.

covariates of particular interest for domain d and ζ_i is a dummy indicating cases featuring a particular design aspect (either polar-point labeling, CAPI, telephone interviewing or two-stage question design). The covariates of special interest, $\boldsymbol{x}_i^{(d)}$, include the gender dummy for all domains d, together with: for $d = H$, the three self-assessed health variables; for $d = I$, the two income variables; for $d = L$, working hours, the gender-hours interaction, the number of children, the interaction of gender and number of children and the childbirth event; for $d = O$, the three self-assessed health variables and the unemployment dummy; and for $d = J$, the hourly wage, working hours and their interactions with gender. Table 5 reports the estimates of $\boldsymbol{\gamma}_d$ and Wald test results for ordinal probit estimates. We also report a variance ratio, defined as:

$$R_d = \frac{Var\left(\boldsymbol{x}_i^{(d)}\hat{\boldsymbol{\gamma}}_d\right)}{Var\left(\boldsymbol{x}_i\hat{\boldsymbol{\beta}}_d\right)} \tag{2}$$

which gives a quantitative measure of the additional variation introduced through design-related instability of the coefficients of variables in $\boldsymbol{x}_i^{(d)}$. It is important to note that the four design contrasts are different splits of largely the same sample, so the four blocks of results in Table 5 are not statistically independent, nor are the columns corresponding to different satisfaction variables.

For satisfaction with health, we find evidence of considerable sensitivity of estimated coefficients to the use of CAPI rather than CASI, CATI rather than CAPI or CASI and two-stage rather than one-stage question design. The effect of these three design aspects is to increase greatly the estimated gradient of satisfaction with health with respect to self-assessed general health. The magnitude of the coefficient instability as measured by the ratio of $Var\left(\boldsymbol{x}_i^{(d)}\hat{\boldsymbol{\gamma}}_d\right)$ relative to $Var\left(\boldsymbol{x}_i\hat{\boldsymbol{\beta}}_d\right)$ is greatest at 21% for CAPI relative to CASI interviewing. This high degree of sensitivity to survey methodology of the conditional relationship between self-assessed health and satisfaction with health is potentially important. Subjective assessments of the quality of life in different health states are used extensively in

health research and our experimental evidence suggests a need for caution in this type of research.

Satisfaction with income displays the largest degree of coefficient instability, but this is only significant for the contrast between one-stage and two-stage question structures, where the variance ratio of 75% is very large indeed. The source of the instability is the coefficient of per capita household income, which is large and significantly different from zero for respondents to the two-stage question but very small and grossly insignificant otherwise. Note that the question asks specifically about satisfaction with household (rather than personal) income, so this coefficient is of central importance for statistical modeling. The behavioural basis for the large effect of question structure on models of income satisfaction is not clear. The survey methods literature (for example, Krosnick and Berent 1993) looks mainly at the influence of question structure on simple summary statistics and is largely silent on the reasons why different types of respondent(members of high- or low-income families) might be affected in different ways by question design. It does not seem possible to give a definite recommendation on which question type is preferable for modeling purposes, but the higher frequency of extreme values produced by the two-stage design seems an undesirable feature.

The relationship between satisfaction with leisure and family structure is significant in the contrasts between CATI and CAPI/CASI and between two-stage and one-stage question design. We find no effect of CATI or the two-stage structure on the estimated influence of the number of children on women's reported satisfaction with leisure, but a large fall for men.[4] Again, it is unclear why this should occur.

The impact of design features on the model of overall satisfaction is moderately large in terms of the variance ratio, but has no clear pattern and is largely insignificant statistically.

[4]This is because the coefficients of the interaction of the design dummy variable with the number of dependent children and with female × the number of dependent children are approximately equal, with opposite signs.

For job satisfaction, there is no direct experimental parallel with the BHPS CAPI/self completion contrast analysed by Conti and Pudney (2010) and the significant tests on individual coefficients are not confirmed by joint Wald tests. The strongest evidence for an impact of survey design is for telephone interviewing contrasted with CAPI/CASI, where the effect is to increase the positive difference between women's and men's general level of expressed job satisfaction and to widen the (negative) male-female difference in the influence of working hours on job satisfaction. This goes some way to confirming Conti and Pudney's (2010) general conclusion that estimates of important parameters in the relationship between job satisfaction and job characteristics like hours and wages are not robust to design features.

Table 5 Tests for structural breaks across treatment groups in satisfaction: coefficients of design interaction terms in ordered probit models

Interaction with...	**Polar-point labeling: shift in marginal effect on satisfaction with...**					**CAPI interviewing: shift in marginal effect on satisfaction with...**				
	Health	Income	Leisure	Overall	Job	Health	Income	Leisure	Overall	Job
Health excellent	0.289			-0.085		-0.147			-0.036	
	(0.235)			(0.240)		(0.263)			(0.238)	
Health fair	-0.171			-0.255		-0.642***			-0.502**	
	(0.206)			(0.218)		(0.221)			(0.246)	
Health poor	-0.234			-0.228		-0.916**			-0.086	
	(0.350)			(0.339)		(0.387)			(0.365)	
Female	0.095	-0.088	-0.341*	0.033	-0.610	0.099	0.146	0.060	0.024	0.310
	(0.145)	(0.123)	(0.175)	(0.131)	(0.701)	(0.152)	(0.134)	(0.200)	(0.145)	(0.762)
Household income		0.000					0.003			
		(0.003)					(0.004)			
Income share		-0.186					0.168			
		(0.209)					(0.232)			
Dependent kids			-0.213					-0.183		
			(0.171)					(0.209)		
Female*no.kids			0.194					0.331		
			(0.195)					(0.232)		
Work hours			-0.007*		-0.007			0.000		-0.013
			(0.004)		(0.012)			(0.005)		(0.012)
Hours*Female			0.002		0.003			-0.003		0.002
			(0.004)		(0.016)			(0.005)		(0.017)
Unemployed				0.712**					-0.022	
				(0.311)					(0.368)	
Wage					-0.032*					0.015
					(0.019)					(0.020)
Wage*female					0.032					-0.031
					(0.026)					(0.030)
Joint P-value	0.516	0.755	0.183	0.281	0.544	0.009**	0.581	0.438	0.523	0.472
Variance ratio	0.042	0.029	0.088	0.131	0.204	0.211	0.047	0.034	0.148	0.331

Significance: * = 10%; ** = 5%; *** = 1%.

Table 5 continued

Interaction with...	**Telephone interviewing: shift in marginal effect on satisfaction with...**					**2-stage question structure: shift in marginal effect on satisfaction with...**				
	Health	Income	Leisure	Overall	Job	Health	Income	Leisure	Overall	Job
Health excellent	0.386**			0.025		0.367*			0.317*	
	(0.187)			(0.162)		(0.192)			(0.164)	
Health fair	-0.188			0.356**		-0.313**			0.145	
	(0.145)			(0.160)		(0.157)			(0.172)	
Health poor	-0.141			0.309		-0.549**			-0.048	
	(0.239)			(0.246)		(0.241)			(0.264)	
Female	0.023	0.141	0.036	0.249**	0.990**	0.191*	0.166	0.195	0.142	0.694
	(0.104)	(0.092)	(0.140)	(0.097)	(0.459)	(0.112)	(0.105)	(0.166)	(0.105)	(0.482)
Household income		0.001					0.015***			
		(0.003)					(0.005)			
Income share		0.011					0.218			
		(0.153)					(0.170)			
Dependent kids			-0.549***					-0.606***		
			(0.155)					(0.129)		
Female*no.kids			0.526***					0.696***		
			(0.170)					(0.149)		
Work hours			-0.001		0.016**			0.001		0.006
			(0.003)		(0.007)			(0.004)		(0.007)
Hours*Female			-0.001		-0.025**			-0.002		-0.005
			(0.004)		(0.010)			(0.005)		(0.011)
Unemployed				0.081					0.384	
				(0.267)					(0.287)	
Wage					-0.006					0.012
					(0.012)					(0.011)
Wage*female					0.004					-0.043**
					(0.018)					(0.020)
Joint P-value	0.124	0.472	0.018**	0.030**	0.209	0.003***	0.011**	0.000***	0.202	0.397
Variance ratio	0.045	0.030	0.016	0.144	0.243	0.100	0.754	0.051	0.127	0.221

Significance: * = 10%; ** = 5%; *** = 1%.

5 Conclusions

The experimental evidence presented here shows quite clearly that the details of survey design – interview mode and question and response scale structure and wording mode – have a significant influence on the distributions of subjective satisfaction data, both unconditional and conditional on other covariates. However, we have found no effect of question placement within the interview. We have confirmed, in general terms, Conti and Pudney's (2010) findings that these design impacts are often significantly different for male and female respondents, so that any analysis of gender differences in satisfaction needs to take account of design features. The contrast between oral interview potentially audible by others (CAPI) and the more private self-interviewing (CASI), and also between the simple one-stage question format and the two-stage format designed for telephone delivery, generate particularly significant differences. These findings suggest that telephone interviewing may be particularly vulnerable to distortion, because the degree of privacy is unknown and the display of response scales via showcards or computer screens is impossible.

For practical purposes, we would recommend the use of a private mode of interview (such as CASI or Self-Completion questionnaires) and a fully-labeled response scale. However, this is a matter of judgement as much as evidence – there is no external measure of 'true' satisfaction which can be used as a validation device, so we have no way of conclusively which of the experimental treatments produces a result closest to the truth.

Perhaps the most important finding of this study is that the impact of these design features is not uniform across types of respondent, so that estimates of statistical models which aim to reveal the relationship between wellbeing and personal circumstances can be affected in important ways by apparently minor features of survey design. The survey methods literature has largely neglected this issue and few of the available theories of respondent behaviour have much to say about the way that characteristics like gender, income and family structure

affect the sensitivity of respondent behaviour to survey design. This is a major challenge for the future, which calls for much closer collaboration between survey design specialists and those who use survey data for statistical modeling.

References

[1] Aquilino, W. S. (1997). Privacy effects on self-reported drug use: interactions with survey mode and respondent characteristics. In Harrison L. and Hughes A. (eds.) *The Validity of Self-Reported Drug Use: Improving the Accuracy of Survey Estimates*, 383-415, Rockville: National Institute on Drug Abuse, NIDA Research Monograph 167.

[2] Clark, A., Frijters, P. and Shields, M. (2008). Relative income, happiness and utility: an explanation for the Easterlin paradox and other puzzles, *Journal of Economic Literature* **46**, 95-144.

[3] Conti, G. and Pudney, S. E. (2010). Survey design and the analysis of satisfaction, *Review of Economics and Statistics* forthcoming.

[4] De Leeuw, E. (1992). Data quality in mail, telephone and face to face surveys

[5] Diener, E., Suh, E. M., Lucas, R. E. and Smith, H. L. (1999). Subjective well-being: three decades of progress, *Psychological Bulletin* **125**, 276-302.

[6] Helgeson, J.G. and Ursic, M.L. (1994). The role of affective and cognitive decision-making processes during questionnaire completion, *Public Opinion Quarterly*, 11(5), 493-510.

[7] Herzog, A.R. and Bachman, J.G. (1981). Effects of questionnaire length on response quality, *Public Opinion Quarterly*, 45, 549-559.

[8] Hunter, J. (2005). Cognitive Test of the 2006 NRFU: Round 1. Washington DC: US Bureau of the Census, Study Series Report (Survey Methodology no.2005-07).

[9] Jenkins, C. R. and Dillman, D. A. (1997). Towards a theory of self-administered questionnaire design. In Lyberg, L., Biemer, P., Collins, M., de Leeuw, E., Dippo, K., Schwarz, N. and Trewin, D. (eds.) *Survey Measurement and Process Quality*. New York: Wiley.

[10] Kahneman, D. and Krueger, A. B. (2006). Developments in the measurement of subjective wellbeing, *Journal of Economic Perspectives* **20**, 3-24.

[11] Kristensen, N. and N. Westergaard-Nielsen (2007). Reliability of job satisfaction measures, *Journal of Happiness Studies*, 8, 273-292.

[12] Krosnick, J. A., and Berent, Matthew K. (1993). Comparisons of party identification and policy preferences: the impact of survey question format, *American Journal of Political Science*, 37(3), 941964.

[13] Krueger, A.B. and D.A. Schkade (2008). The reliability of subjective well-being measures, *Journal of Public Economics*, 92(8-9), 1833-1845.

[14] Pudney, S. E. (2010). The Understanding Society Innovation Panel: Notes on the construction of gross annual income variables for wave 2. Mimeo, University of Essex.

[15] Schuman, H. and Presser, S. (1981). *Questions and Answers in Attitude Surveys: Experiments in Question Form, Wording and Context*. New York: Academic Press.

[16] Schhwarz, N., Strackk, F. and Mai, H. (1991). Assimilation and contrast effects in part-whole question sequences: a conversational logic analysis, *Public Opinion Quarterly*, 55, 3-23.

[17] Smith, T.W. (1979). Happiness: Time Trends, Seasonal Variations, Intersurvey Differences, and Other Mysteries, *Social Psychology Quarterly*, 42(1), 18-30.

[18] Tourangeau, R., K.A. Rasinski and N. Bradburn (1991), "Measuring happiness in surveys: a test of the subtraction hypothesis", *Public Opinion Quarterly*, **55**, 255-266.

[19] Tourangeau, R. (1999). Context effects on answers to attitude questions, in Sirken, M. G., Herrmann, D. J., Schechter, S., Schwarz, N., Tanur, J. M. and Tourangeau, R. (eds.) *Cognition and Survey Research*. New York: Wiley.

[20] Weng, L.-J. (2004). Impact of the number of response categories and anchor labels on coefficient alpha and test-retest reliability, *Educational and Psychological Measurement*, 64, 956-972.

[21] Van Praag, B. and Ferrer-i-Carbonell, A. (2004). *Happiness Quantified. A Satisfaction Calculus Approach*. Oxford: Oxford University Press.

Appendix: Additional tables

Table A1 Covariate sample means (full sample, $n = 1782$)

Covariate	Mean	Covariate	Mean
Excellent health	0.149	Household income per head (£'000 p.a.)[a]	14.60
Fair health	0.145	Personal share in household income[a]	0.603
Poor health	0.065	Degree-level qualification	0.221
At least one disability	0.311	No qualifications	0.180
Diagnosed serious disease	0.431	Managerial/professional	0.240
Age/10	4.95	Unskilled worker	0.132
Household size	2.68	Employed or self-employed	0.551
Single	0.127	Unemployed	0.052
Divorced, separated or widowed	0.052	Retired	0.263
Cohabiting	0.084	Private sector worker	0.304
Absent partner	0.347	Pension scheme member	0.246
No. of dependent children	0.281	Hours of work[b]	37.4
New birth in last year	0.043	Hourly wage[b]	11.08
Non-white	0.062		

[a] See Pudney (2010) for explanation of the method of constructing IP2 income variables; [b] Mean computed from positive sample values.

Table A2 Ordered probit models

Covariate	Health		Income		Leisure		Life overall		Job	
	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male
Health excellent	0.950***	0.915***	0.245**	0.406***	0.149	-0.008	0.378***	0.230*	0.335***	0.236*
	(0.129)	(0.140)	(0.103)	(0.113)	(0.104)	(0.113)	(0.107)	(0.121)	(0.123)	(0.134)
Health fair	-1.002***	-0.776***	-0.416***	-0.266**	-0.381***	-0.101	-0.558***	-0.403***	-0.300*	-0.545***
	(0.107)	(0.123)	(0.107)	(0.126)	(0.107)	(0.125)	(0.111)	(0.124)	(0.182)	(0.187)
Health poor	-1.958***	-2.176***	-0.659***	-0.680***	-0.422**	-0.594***	-0.847***	-0.894***	0.141	0.331
	(0.179)	(0.230)	(0.188)	(0.183)	(0.193)	(0.195)	(0.180)	(0.193)	(0.297)	(0.320)
Disabled	-0.303***	-0.346***								
	(0.090)	(0.098)								
Diagnosis	0.413***	0.291***								
	(0.085)	(0.087)								
Age	-0.184	-0.184	-0.546***	-0.315**	-0.062	-0.037	-0.314**	-0.386**	-0.380	-0.794***
	(0.141)	(0.167)	(0.139)	(0.159)	(0.142)	(0.170)	(0.153)	(0.168)	(0.298)	(0.272)
Age2	0.021	0.022	0.068***	0.037**	0.008	0.013	0.029*	0.045***	0.055	0.092***
	(0.014)	(0.017)	(0.014)	(0.016)	(0.015)	(0.018)	(0.016)	(0.017)	(0.035)	(0.030)
Hh size	0.012	0.014	-0.001	-0.075*	-0.001	-0.078	-0.043	-0.049	0.079	-0.017
	(0.043)	(0.051)	(0.049)	(0.042)	(0.044)	(0.048)	(0.047)	(0.047)	(0.064)	(0.055)
Single	0.068	0.094	-0.033	0.518**	-0.262	0.321	0.241	0.314	-0.014	0.370
	(0.191)	(0.234)	(0.198)	(0.225)	(0.198)	(0.212)	(0.200)	(0.239)	(0.285)	(0.289)
Divorced/widowed	-0.037	0.184	-0.365**	0.104	-0.299*	0.111	-0.004	-0.196	-0.180	0.006
	(0.183)	(0.333)	(0.169)	(0.340)	(0.173)	(0.308)	(0.182)	(0.327)	(0.269)	(0.410)
Cohabiting	-0.158	-0.145	-0.215*	-0.149	-0.138	-0.316**	-0.197	-0.072	-0.147	-0.176
	(0.133)	(0.147)	(0.125)	(0.144)	(0.135)	(0.124)	(0.132)	(0.124)	(0.142)	(0.168)
Absent partner	0.053	-0.140	-0.117	-0.179	0.395***	-0.054	-0.311**	-0.303*	-0.050	-0.421*
	(0.135)	(0.165)	(0.140)	(0.154)	(0.135)	(0.165)	(0.137)	(0.164)	(0.214)	(0.216)
No. dep. children	0.037	-0.015	-0.032	-0.064	-0.146**	-0.010	0.006	0.121	0.036	0.082
	(0.061)	(0.193)	(0.060)	(0.292)	(0.058)	(0.176)	(0.060)	(0.237)	(0.085)	(0.218)
Recent birth	-0.028	-0.052	-0.171	0.041	-0.399**	0.010	-0.049	-0.045	-0.344	-0.239
	(0.166)	(0.209)	(0.162)	(0.199)	(0.172)	(0.186)	(0.196)	(0.181)	(0.213)	(0.189)
Non-white	-0.121	-0.032	-0.192	-0.132	-0.301**	-0.310**	-0.199	-0.041	-0.202	-0.072
	(0.129)	(0.164)	(0.147)	(0.154)	(0.143)	(0.150)	(0.150)	(0.181)	(0.210)	(0.176)

Table A2 continued

Covariate	Health		Income		Leisure		Life overall		Job	
	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male
In work	0.031	0.314	-0.109	0.175	0.209	-0.083	0.130	0.158		
	(0.152)	(0.218)	(0.148)	(0.243)	(0.143)	(0.230)	(0.149)	(0.217)		
Unemployed	-0.262	-0.089	-0.561***	-0.747***	-0.315	-0.218	-0.327	-0.564**		
	(0.184)	(0.233)	(0.208)	(0.266)	(0.192)	(0.235)	(0.215)	(0.223)		
Retired	0.182	0.227	-0.281*	0.043	0.762***	0.597**	0.379**	0.423*		
	(0.156)	(0.236)	(0.158)	(0.239)	(0.177)	(0.242)	(0.162)	(0.224)		
Hh income p.c.	-0.001	0.002	0.002	-0.001	0.001	-0.002	0.000	-0.002	0.003*	0.003**
	(0.001)	(0.002)	(0.002)	(0.002)	(0.001)	(0.003)	(0.001)	(0.001)	(0.002)	(0.002)
incshare	-0.062	0.103	-0.311**	0.137	-0.363***	0.013	-0.372***	-0.147	0.339*	0.620**
	(0.129)	(0.182)	(0.128)	(0.172)	(0.129)	(0.193)	(0.129)	(0.186)	(0.186)	(0.268)
Hourly wage	0.011	-0.001	0.004	0.026***	0.008	0.012*	0.011	0.007	-0.002	0.013*
	(0.009)	(0.006)	(0.010)	(0.006)	(0.007)	(0.007)	(0.008)	(0.005)	(0.008)	(0.007)
Work hours	0.001	-0.007**	0.003	-0.005	-0.020***	-0.010***	-0.004	0.000	-0.009**	0.001
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.004)	(0.004)	(0.004)
Degree	-0.104	-0.160	0.145	0.032	-0.048	-0.083	-0.266***	-0.152	-0.139	-0.278**
	(0.105)	(0.102)	(0.101)	(0.100)	(0.098)	(0.098)	(0.094)	(0.099)	(0.121)	(0.122)
No quals	0.151	0.248*	-0.059	0.054	0.031	-0.056	0.292***	0.264**	0.392*	0.129
	(0.102)	(0.128)	(0.116)	(0.118)	(0.110)	(0.126)	(0.112)	(0.119)	(0.213)	(0.192)
Manager/prof	0.026	0.087	0.047	0.122	0.002	-0.055	0.071	0.151	0.082	0.193
	(0.099)	(0.096)	(0.096)	(0.087)	(0.098)	(0.096)	(0.098)	(0.098)	(0.123)	(0.125)
Unskilled	0.156	0.184	-0.028	0.008	0.003	-0.019	0.066	0.052	-0.003	-0.087
	(0.118)	(0.133)	(0.117)	(0.136)	(0.113)	(0.129)	(0.118)	(0.126)	(0.193)	(0.183)
Private sector	0.046	0.114	-0.013	-0.226**	0.158*	-0.158	0.047	-0.040	-0.242**	-0.406***
	(0.101)	(0.108)	(0.086)	(0.101)	(0.091)	(0.099)	(0.091)	(0.098)	(0.103)	(0.106)
In pension scheme	0.004	0.188	-0.017	0.175*	0.129	0.052	0.101	0.155	0.127	0.117
	(0.114)	(0.115)	(0.111)	(0.102)	(0.101)	(0.104)	(0.101)	(0.102)	(0.112)	(0.118)

Table A2 continued

Covariate	Health		Income		Leisure		Life overall		Job	
	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male
Group B	-0.151	-0.383*	-0.381**	0.036	-0.256*	0.232	-0.289*	-0.108	-0.198	0.182
	(0.148)	(0.210)	(0.156)	(0.173)	(0.145)	(0.180)	(0.154)	(0.188)	(0.206)	(0.195)
Group C	0.164	0.452**	-0.203	0.013	0.249*	0.229	0.334*	0.742***	0.428*	0.298
	(0.178)	(0.218)	(0.172)	(0.241)	(0.145)	(0.284)	(0.171)	(0.276)	(0.234)	(0.334)
Group D (tel)	0.153	-0.183	-0.004	-0.237*	0.140	0.110	0.229*	0.155	0.469***	0.455***
	(0.140)	(0.149)	(0.127)	(0.141)	(0.125)	(0.139)	(0.132)	(0.135)	(0.142)	(0.166)
Group D (F2F)	0.442*	-0.232	-0.004	-0.231	0.127	0.314	0.327	0.042	0.212	0.275
	(0.237)	(0.203)	(0.194)	(0.213)	(0.214)	(0.226)	(0.214)	(0.229)	(0.315)	(0.263)
Group E	-0.401	-0.323	-0.294	-0.048	-0.145	-0.526***	-0.187	-0.188	0.430	-0.347
	(0.281)	(0.253)	(0.214)	(0.243)	(0.218)	(0.201)	(0.218)	(0.212)	(0.268)	(0.254)
Group F	-0.003	-0.056	-0.177	-0.192	-0.255**	0.062	0.081	-0.020	0.069	-0.005
	(0.131)	(0.132)	(0.116)	(0.122)	(0.118)	(0.125)	(0.120)	(0.115)	(0.127)	(0.124)
Group H (tel)	0.195	0.141	-0.164	0.087	0.167	0.107	0.463***	0.138		
	(0.162)	(0.171)	(0.155)	(0.179)	(0.150)	(0.180)	(0.166)	(0.165)		
Group H (F2F)	-0.088	0.054	-0.138	0.116	-0.017	0.080	0.019	0.387		
	(0.200)	(0.293)	(0.204)	(0.254)	(0.193)	(0.300)	(0.184)	(0.242)		
Group I	0.178	-0.042	-0.269	0.129	-0.291	0.126	-0.236	-0.103		
	(0.250)	(0.294)	(0.219)	(0.225)	(0.210)	(0.263)	(0.230)	(0.256)		
Group J	0.143	0.013	-0.067	-0.085	-0.055	0.200	0.017	0.109		
	(0.140)	(0.152)	(0.136)	(0.144)	(0.139)	(0.148)	(0.134)	(0.132)		
Threshold_1	-2.466***	-2.805***	-3.210***	-2.679***	-2.078***	-2.265***	-3.815***	-3.371***	-2.524***	-3.690***
	(0.449)	(0.567)	(0.463)	(0.535)	(0.468)	(0.546)	(0.508)	(0.520)	(0.712)	(0.665)
Threshold_2	-1.959***	-2.446***	-2.833***	-2.189***	-1.444***	-1.736***	-3.308***	-2.988***	-2.054***	-3.372***
	(0.449)	(0.560)	(0.461)	(0.531)	(0.462)	(0.541)	(0.498)	(0.518)	(0.723)	(0.651)
Threshold_3	-1.313***	-1.882***	-2.359***	-1.714***	-0.879*	-1.262**	-2.981***	-2.640***	-1.641**	-2.834***
	(0.450)	(0.556)	(0.464)	(0.527)	(0.460)	(0.538)	(0.494)	(0.513)	(0.720)	(0.649)
Threshold_4	-0.920**	-1.230**	-1.782***	-1.177**	-0.474	-0.785	-2.412***	-2.010***	-1.205*	-2.282***
	(0.448)	(0.550)	(0.461)	(0.527)	(0.461)	(0.537)	(0.487)	(0.506)	(0.720)	(0.649)
Threshold_5	-0.191	-0.478	-1.231***	-0.477	-0.001	-0.278	-1.740***	-1.346***	-0.585	-1.687***
	(0.447)	(0.551)	(0.460)	(0.525)	(0.461)	(0.535)	(0.485)	(0.503)	(0.716)	(0.650)
Threshold_6	0.955**	0.663	-0.346	0.416	0.778*	0.384	-0.549	-0.149	0.383	-0.733
	(0.445)	(0.550)	(0.458)	(0.526)	(0.462)	(0.532)	(0.482)	(0.497)	(0.714)	(0.650)

Table A3 Regression models

Covariate	Health		Income		Leisure		Life overall		Job	
	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male
Health excellent	0.742***	0.696***	0.335**	0.453***	0.170	-0.043	0.315***	0.188	0.374**	0.227
	(0.107)	(0.110)	(0.146)	(0.147)	(0.151)	(0.172)	(0.104)	(0.133)	(0.145)	(0.161)
Health fair	-1.360***	-0.964***	-0.634***	-0.445**	-0.515***	-0.146	-0.668***	-0.446***	-0.479*	-0.875***
	(0.148)	(0.151)	(0.166)	(0.189)	(0.157)	(0.170)	(0.135)	(0.148)	(0.270)	(0.274)
Health poor	-2.808***	-2.953***	-1.011***	-1.050***	-0.594**	-0.824***	-1.075***	-1.128***	0.249	0.387
	(0.211)	(0.256)	(0.292)	(0.274)	(0.278)	(0.285)	(0.239)	(0.260)	(0.418)	(0.353)
Disabled	-0.419***	-0.353***								
	(0.117)	(0.115)								
Diagnosis	0.435***	0.316***								
	(0.097)	(0.091)								
Age	-0.208	-0.148	-0.782***	-0.474**	-0.204	0.074	-0.376**	-0.385**	-0.349	-1.178***
	(0.171)	(0.186)	(0.204)	(0.234)	(0.198)	(0.223)	(0.178)	(0.188)	(0.386)	(0.349)
Age2	0.027	0.021	0.098***	0.055**	0.022	0.005	0.036**	0.045**	0.056	0.136***
	(0.017)	(0.019)	(0.020)	(0.023)	(0.019)	(0.021)	(0.018)	(0.019)	(0.044)	(0.039)
Hh size	0.030	0.014	0.005	-0.103	-0.016	-0.128*	-0.021	-0.055	0.090	-0.003
	(0.049)	(0.053)	(0.074)	(0.063)	(0.071)	(0.077)	(0.052)	(0.055)	(0.089)	(0.075)
Single	0.113	0.186	-0.023	0.772**	-0.329	0.639**	0.287	0.485*	0.004	0.347
	(0.225)	(0.265)	(0.304)	(0.330)	(0.300)	(0.317)	(0.243)	(0.280)	(0.396)	(0.375)
Divorced/widowed	0.054	0.030	-0.536**	0.124	-0.361	0.121	0.026	-0.326	-0.204	-0.095
	(0.211)	(0.373)	(0.265)	(0.490)	(0.245)	(0.437)	(0.217)	(0.448)	(0.363)	(0.568)
Cohabiting	-0.134	-0.141	-0.319	-0.260	-0.288	-0.525**	-0.208	-0.015	-0.126	-0.139
	(0.160)	(0.168)	(0.196)	(0.217)	(0.218)	(0.209)	(0.153)	(0.151)	(0.207)	(0.235)
Absent partner	0.004	-0.087	-0.203	-0.266	0.436**	-0.121	-0.369**	-0.327*	-0.074	-0.401
	(0.160)	(0.179)	(0.212)	(0.220)	(0.187)	(0.223)	(0.154)	(0.186)	(0.287)	(0.286)
No. dep. children	0.019	-0.035	-0.040	-0.080	-0.220**	0.056	0.016	0.162	0.072	0.216
	(0.069)	(0.216)	(0.091)	(0.446)	(0.092)	(0.300)	(0.066)	(0.332)	(0.111)	(0.287)
Recent birth	-0.024	0.031	-0.269	0.049	-0.632**	0.033	-0.080	0.011	-0.352	-0.257
	(0.198)	(0.211)	(0.259)	(0.290)	(0.291)	(0.316)	(0.231)	(0.233)	(0.305)	(0.266)
Non-white	-0.090	-0.027	-0.307	-0.205	-0.507**	-0.529**	-0.254	-0.168	-0.263	-0.018
	(0.162)	(0.177)	(0.226)	(0.225)	(0.238)	(0.251)	(0.182)	(0.219)	(0.303)	(0.249)

Table A3 continued

Covariate	Health		Income		Leisure		Life overall		Job	
	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male
In work	0.025	0.345	-0.164	0.251	0.415*	-0.014	0.241	0.256		
	(0.174)	(0.237)	(0.222)	(0.344)	(0.222)	(0.355)	(0.170)	(0.266)		
Unemployed	-0.301	-0.147	-0.858***	-1.211***	-0.426	-0.366	-0.360	-0.785**		
	(0.233)	(0.282)	(0.318)	(0.387)	(0.304)	(0.366)	(0.267)	(0.317)		
Retired	0.220	0.243	-0.357	0.088	0.982***	0.671**	0.490**	0.542**		
	(0.189)	(0.268)	(0.233)	(0.346)	(0.235)	(0.336)	(0.192)	(0.272)		
Hh income p.c.	-0.001	0.002	0.004	-0.001	0.001	-0.002	0.001	-0.002	0.003***	0.005**
	(0.001)	(0.001)	(0.003)	(0.003)	(0.001)	(0.004)	(0.002)	(0.002)	(0.001)	(0.002)
incshare	-0.025	0.094	-0.454**	0.208	-0.507***	0.015	-0.393***	-0.221	0.394	0.640*
	(0.150)	(0.194)	(0.198)	(0.252)	(0.192)	(0.273)	(0.140)	(0.217)	(0.241)	(0.351)
Hourly wage	0.013	-0.001	0.004	0.033***	0.011	0.019*	0.012	0.009	-0.003	0.017**
	(0.010)	(0.006)	(0.015)	(0.008)	(0.010)	(0.010)	(0.009)	(0.006)	(0.010)	(0.008)
Work hours	0.003	-0.007**	0.006	-0.006	-0.032***	-0.017***	-0.003	0.000	-0.011*	0.004
	(0.004)	(0.003)	(0.005)	(0.005)	(0.005)	(0.006)	(0.004)	(0.004)	(0.006)	(0.006)
Degree	-0.143	-0.164	0.216	0.025	-0.137	-0.132	-0.296***	-0.131	-0.203	-0.366**
	(0.124)	(0.105)	(0.152)	(0.143)	(0.153)	(0.151)	(0.112)	(0.111)	(0.161)	(0.168)
No quals	0.153	0.242*	-0.098	0.063	0.085	0.044	0.273**	0.263**	0.439	0.218
	(0.121)	(0.135)	(0.171)	(0.169)	(0.144)	(0.168)	(0.117)	(0.122)	(0.269)	(0.251)
Manager/prof	0.043	0.081	0.053	0.194	0.012	-0.127	0.079	0.127	0.134	0.239
	(0.114)	(0.102)	(0.146)	(0.126)	(0.148)	(0.139)	(0.106)	(0.109)	(0.161)	(0.168)
Unskilled	0.182	0.102	-0.052	-0.032	-0.017	-0.032	0.076	0.030	-0.025	-0.139
	(0.136)	(0.137)	(0.177)	(0.202)	(0.158)	(0.191)	(0.132)	(0.148)	(0.259)	(0.255)
Private sector	0.028	0.096	-0.010	-0.310**	0.268*	-0.279*	0.048	-0.034	-0.297**	-0.507***
	(0.108)	(0.113)	(0.133)	(0.141)	(0.147)	(0.164)	(0.097)	(0.109)	(0.137)	(0.137)
Pension scheme	0.024	0.225*	0.011	0.333**	0.286*	0.142	0.151	0.259**	0.291*	0.266*
	(0.124)	(0.117)	(0.176)	(0.145)	(0.167)	(0.170)	(0.111)	(0.112)	(0.156)	(0.153)

Table A3 continued

Covariate	Health		Income		Leisure		Life overall		Job	
	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male
Group B	-0.138	-0.481**	-0.565**	0.060	-0.382*	0.239	-0.323*	-0.240	-0.384	0.216
	(0.179)	(0.242)	(0.241)	(0.254)	(0.226)	(0.267)	(0.189)	(0.218)	(0.307)	(0.258)
Group C	0.133	0.433**	-0.306	0.016	0.332	0.342	0.388**	0.617**	0.575**	0.343
	(0.221)	(0.205)	(0.276)	(0.320)	(0.222)	(0.406)	(0.180)	(0.272)	(0.273)	(0.444)
Group D (tel)	0.149	-0.302*	-0.039	-0.361*	0.143	0.188	0.201	0.095	0.481***	0.511**
	(0.167)	(0.166)	(0.189)	(0.207)	(0.183)	(0.210)	(0.150)	(0.154)	(0.175)	(0.208)
Group D (F2F)	0.310	-0.324	-0.057	-0.374	0.103	0.423	0.289	-0.055	0.085	0.326
	(0.258)	(0.229)	(0.282)	(0.311)	(0.302)	(0.320)	(0.220)	(0.261)	(0.436)	(0.347)
Group E	-0.536	-0.384	-0.551*	-0.127	-0.308	-0.826**	-0.240	-0.254	0.467	-0.578
	(0.364)	(0.290)	(0.325)	(0.338)	(0.325)	(0.320)	(0.268)	(0.269)	(0.363)	(0.371)
Group F	0.020	-0.082	-0.213	-0.251	-0.413**	0.118	0.080	-0.035	0.111	0.021
	(0.161)	(0.147)	(0.176)	(0.179)	(0.175)	(0.192)	(0.139)	(0.132)	(0.168)	(0.172)
Group H (tel)	0.147	0.109	-0.251	0.089	0.139	0.144	0.435**	0.103		
	(0.185)	(0.184)	(0.235)	(0.254)	(0.221)	(0.269)	(0.181)	(0.183)		
Group H (F2F)	-0.096	-0.086	-0.206	0.070	-0.058	0.056	0.013	0.235		
	(0.245)	(0.288)	(0.312)	(0.355)	(0.298)	(0.441)	(0.208)	(0.262)		
Group I	0.109	-0.169	-0.391	0.208	-0.476	0.121	-0.320	-0.163		
	(0.274)	(0.331)	(0.336)	(0.325)	(0.302)	(0.369)	(0.285)	(0.304)		
Group J	0.196	-0.052	-0.147	-0.138	-0.201	0.247	-0.017	0.047		
	(0.168)	(0.166)	(0.200)	(0.209)	(0.202)	(0.222)	(0.157)	(0.148)		
Intercept	5.351***	5.599***	6.905***	6.030***	5.561***	5.361***	6.991***	6.582***	5.582***	7.348***
	(0.536)	(0.613)	(0.687)	(0.762)	(0.698)	(0.785)	(0.542)	(0.560)	(0.945)	(0.828)

Table A4 Probit models for "mostly" or "completely" satisfied

Covariate	Health		Income		Leisure		Life overall		Job	
	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male
Health excellent	0.859***	1.008***	0.238*	0.583***	0.197	-0.111	0.247*	0.299**	0.260*	0.149
	(0.164)	(0.183)	(0.127)	(0.137)	(0.130)	(0.146)	(0.138)	(0.149)	(0.155)	(0.162)
Health fair	-1.149***	-0.899***	-0.340***	-0.202	-0.391***	-0.020	-0.638***	-0.356**	-0.219	-0.527**
	(0.155)	(0.154)	(0.130)	(0.146)	(0.134)	(0.152)	(0.128)	(0.144)	(0.204)	(0.216)
Health poor	-1.557***	-1.843***	-0.664***	-0.740***	-0.261	-0.694***	-0.695***	-0.859***	0.317	-0.036
	(0.293)	(0.321)	(0.217)	(0.250)	(0.207)	(0.219)	(0.194)	(0.220)	(0.389)	(0.435)
Disabled	-0.367***	-0.340***								
	(0.117)	(0.123)								
Diagnosis	0.390***	0.236**								
	(0.108)	(0.116)								
Age	-0.178	-0.168	-0.473***	-0.072	0.126	0.056	-0.248	-0.481**	-0.581*	-0.542
	(0.189)	(0.204)	(0.172)	(0.191)	(0.170)	(0.219)	(0.177)	(0.199)	(0.347)	(0.361)
Age2	0.017	0.021	0.061***	0.012	-0.010	0.003	0.024	0.047**	0.081**	0.069*
	(0.019)	(0.021)	(0.017)	(0.019)	(0.018)	(0.022)	(0.018)	(0.020)	(0.040)	(0.041)
Hh size	-0.011	0.065	0.004	-0.140**	-0.039	-0.017	-0.084	-0.075	0.056	-0.028
	(0.063)	(0.061)	(0.056)	(0.058)	(0.057)	(0.058)	(0.063)	(0.059)	(0.079)	(0.067)
Single	-0.179	0.047	-0.055	0.617**	-0.156	0.030	0.275	0.446	-0.020	0.517
	(0.246)	(0.302)	(0.242)	(0.278)	(0.232)	(0.282)	(0.244)	(0.296)	(0.339)	(0.355)
Divorced/widowed	-0.072	0.525	-0.509**	-0.047	-0.515**	-0.318	0.082	-0.026	-0.021	-0.120
	(0.219)	(0.385)	(0.219)	(0.402)	(0.217)	(0.391)	(0.219)	(0.379)	(0.321)	(0.557)
Cohabiting	-0.296	-0.251	-0.383**	-0.120	-0.146	-0.588***	-0.255	0.007	-0.099	-0.110
	(0.182)	(0.201)	(0.174)	(0.194)	(0.170)	(0.211)	(0.166)	(0.186)	(0.204)	(0.208)
Absent partner	0.085	-0.128	-0.093	-0.167	0.474***	0.111	-0.299*	-0.539***	-0.071	-0.386
	(0.179)	(0.211)	(0.168)	(0.191)	(0.170)	(0.205)	(0.171)	(0.197)	(0.257)	(0.278)
No. dep. children	0.024	-0.347	-0.011	0.176	-0.098	-0.057	0.025	0.216	0.111	0.565
	(0.080)	(0.299)	(0.075)	(0.351)	(0.074)	(0.292)	(0.075)	(0.317)	(0.107)	(0.370)
Recent birth	-0.076	-0.101	-0.384*	0.002	-0.482**	0.120	0.159	0.118	-0.270	-0.241
	(0.226)	(0.270)	(0.229)	(0.249)	(0.240)	(0.238)	(0.244)	(0.238)	(0.281)	(0.257)
Non-white	-0.247	0.112	-0.224	-0.157	-0.252	-0.373*	-0.381**	-0.097	-0.216	0.002
	(0.196)	(0.218)	(0.190)	(0.200)	(0.195)	(0.196)	(0.194)	(0.211)	(0.245)	(0.247)

Table A4 continued

Covariate	Health		Income		Leisure		Life overall		Job	
	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male
In work	0.004	0.444	-0.038	0.230	0.342*	0.148	0.036	0.114		
	(0.187)	(0.286)	(0.184)	(0.267)	(0.180)	(0.275)	(0.181)	(0.264)		
Unemployed	-0.558**	-0.233	-0.581**	-0.200	-0.412*	-0.006	-0.452*	-0.554**		
	(0.284)	(0.299)	(0.282)	(0.284)	(0.226)	(0.265)	(0.240)	(0.277)		
Retired	0.263	0.323	-0.255	0.185	0.772***	0.792***	0.245	0.590**		
	(0.220)	(0.303)	(0.193)	(0.275)	(0.198)	(0.298)	(0.188)	(0.278)		
Hh income p.c.	-0.002	0.001	0.000	-0.000	0.001	-0.000	-0.000	-0.000	0.004*	0.008**
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.004)
incshare	-0.051	-0.114	-0.279*	0.084	-0.379**	-0.032	-0.460***	0.025	0.158	0.906***
	(0.171)	(0.235)	(0.160)	(0.224)	(0.161)	(0.229)	(0.166)	(0.229)	(0.232)	(0.332)
Hourly wage	0.005	-0.001	0.008	0.032***	-0.008	0.005	0.014	0.015*	-0.010	0.010
	(0.011)	(0.009)	(0.009)	(0.009)	(0.010)	(0.009)	(0.012)	(0.009)	(0.010)	(0.010)
Work hours	0.002	-0.009**	0.003	-0.009**	-0.023***	-0.015***	-0.006	-0.002	-0.009**	-0.001
	(0.005)	(0.004)	(0.004)	(0.004)	(0.005)	(0.004)	(0.005)	(0.004)	(0.005)	(0.005)
Degree	-0.095	-0.153	0.091	0.079	-0.093	-0.028	-0.272**	-0.140	-0.181	-0.332**
	(0.135)	(0.137)	(0.126)	(0.131)	(0.128)	(0.130)	(0.130)	(0.131)	(0.149)	(0.159)
No quals	-0.089	0.355**	0.018	-0.014	0.018	-0.033	0.262*	0.157	0.131	0.184
	(0.133)	(0.163)	(0.132)	(0.146)	(0.136)	(0.162)	(0.138)	(0.150)	(0.236)	(0.249)
Manager/prof	-0.026	0.016	0.092	0.177	-0.145	-0.077	-0.130	0.147	0.078	0.313**
	(0.133)	(0.129)	(0.125)	(0.121)	(0.129)	(0.124)	(0.131)	(0.123)	(0.158)	(0.159)
Unskilled	0.121	0.053	-0.235	0.134	-0.046	-0.041	-0.066	-0.006	-0.093	-0.039
	(0.147)	(0.163)	(0.143)	(0.154)	(0.139)	(0.162)	(0.142)	(0.152)	(0.215)	(0.215)
Private sector	0.026	0.070	0.024	-0.115	0.217*	-0.143	0.214	-0.070	-0.149	-0.476***
	(0.131)	(0.143)	(0.120)	(0.136)	(0.128)	(0.136)	(0.131)	(0.134)	(0.130)	(0.141)
Pension scheme	-0.002	0.289*	-0.206	0.183	0.255*	0.067	0.333**	0.276*	0.169	0.321**
	(0.150)	(0.157)	(0.139)	(0.144)	(0.143)	(0.140)	(0.153)	(0.143)	(0.145)	(0.143)

Table A4 continued

Covariate	Health		Income		Leisure		Life overall		Job	
	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male
Group B	-0.615***	-0.803***	-0.714***	0.043	-0.394**	0.169	-0.542***	-0.363	-0.320	0.151
	(0.204)	(0.255)	(0.196)	(0.230)	(0.201)	(0.230)	(0.205)	(0.237)	(0.219)	(0.240)
Group C	-0.159	0.912***	-0.303	-0.292	-0.078	0.318	0.373	1.129***	0.214	0.505
	(0.263)	(0.322)	(0.250)	(0.309)	(0.250)	(0.342)	(0.298)	(0.360)	(0.301)	(0.397)
Group D (tel)	-0.229	-0.375**	-0.107	-0.284	0.086	0.307*	-0.013	0.138	0.370**	0.348*
	(0.183)	(0.191)	(0.157)	(0.182)	(0.170)	(0.185)	(0.170)	(0.183)	(0.174)	(0.189)
Group D (F2F)	0.121	-0.572**	-0.204	-0.294	0.091	0.492	0.183	0.003	0.515	0.413
	(0.283)	(0.259)	(0.239)	(0.269)	(0.270)	(0.308)	(0.272)	(0.275)	(0.354)	(0.326)
Group E	-0.807***	-0.533*	-0.538**	-0.281	-0.428	-0.919***	-0.695***	-0.400	0.638**	-0.651**
	(0.283)	(0.303)	(0.249)	(0.261)	(0.268)	(0.300)	(0.261)	(0.277)	(0.303)	(0.301)
Group F	-0.537***	-0.447**	-0.484***	-0.515***	-0.596***	-0.022	-0.265*	-0.087	-0.193	-0.145
	(0.170)	(0.179)	(0.155)	(0.172)	(0.165)	(0.174)	(0.160)	(0.164)	(0.164)	(0.167)
Group H (tel)	-0.309	0.075	-0.271	0.056	0.075	0.210	0.474**	0.083		
	(0.208)	(0.236)	(0.199)	(0.228)	(0.203)	(0.218)	(0.241)	(0.226)		
Group H (F2F)	-0.355	-0.370	-0.276	0.148	-0.037	0.320	-0.239	0.220		
	(0.302)	(0.357)	(0.256)	(0.282)	(0.261)	(0.293)	(0.242)	(0.281)		
Group I	0.017	-0.416	-0.394	-0.014	-0.642**	-0.352	-0.327	-0.332		
	(0.287)	(0.360)	(0.265)	(0.306)	(0.279)	(0.352)	(0.255)	(0.327)		
Group J	-0.590***	-0.545***	-0.439***	-0.324*	-0.349**	0.074	-0.456***	-0.016		
	(0.186)	(0.204)	(0.170)	(0.193)	(0.176)	(0.195)	(0.171)	(0.186)		
Intercept	0.917	0.716	1.238**	0.159	-0.237	-0.290	2.012***	2.014***	1.174	0.835
	(0.617)	(0.686)	(0.575)	(0.641)	(0.556)	(0.699)	(0.600)	(0.643)	(0.860)	(0.873)