CWiPP Working Paper Series

No.13

OPERATIONALISING EQUIVALENT CONSUMPTION THROUGH STATED PREFERENCES: A PILOT STUDY IN TWO PARTS

Ignacio Abasolo, Departamento de Economía Aplicada y Métodos Cuantitativos, Universidad de La Laguna

Clara Sandelind, Department of Politics, University of Sheffield

Erik Schokkaert, Department of Economics, University of Leuven and CORE, Université catholique de Louvan

Katherine Stevens, School of Health and Related Research, University of Sheffield

Aki Tsuchiya*, School of Health and Related Research, and Department of Economics, University of Sheffield a.tsuchiya@sheffield.ac.uk

*   Corresponding author

www.sheffield.ac.uk/cwipp/working-papers

September 2018

Operationalising Equivalent Consumption through Stated Preferences:
A Pilot Study in Two Parts

Ignacio Abasolo[1], Clara Sandelind[2], Erik Schokkaert[3], Katherine Stevens[4], Aki Tsuchiya[5*]

1. Departamento de Economía Aplicada y Métodos Cuantitativos, Universidad de La Laguna
2. Department of Politics, University of Sheffield
3. Department of Economics, University of Leuven and CORE, Université catholique de Louvan
4. School of Health and Related Research, University of Sheffield
5. School of Health and Related Research, and Department of Economics, University of Sheffield
*   Corresponding author (a.tsuchiya@sheffield.ac.uk)

ABSTRACT

Equivalent consumption (also known as equivalent income) is a preference-based, cardinal, and interpersonally comparable measure of individual wellbeing. This paper reports on a pair of exploratory studies that aimed to estimate the parameters of an individual utility function, to operationalise equivalent consumption. First, a discrete choice experiment was designed to model individual preferences, and pilot data were collected from a convenience sample on-line (n=52). Second, the same survey was conducted in face-to-face interviews with a second sample to qualitatively analyse the way people attempt the survey. The results suggest that, while the choice experiment data can be modelled by pooling across respondents, it is difficult to estimate individual-level models. Furthermore, the qualitative data suggest that respondents struggle to interpret some of the elements of the choice task. The pilot study makes clear the central challenges in taking into account preference heterogeneity when trying to operationalise the concept of equivalent consumption as a measure of individual wellbeing through stated preferences.

KEY WORDS

Equivalent consumption, individual utility function, wellbeing, personal and social perspectives

1. INTRODUCTION

Human wellbeing can be captured at the *population* level (e.g. GDP per capita) or at the *individual* level (e.g. individual disposable income). Individual wellbeing can be captured as an *objective* measure (e.g. disposable income) or as a *subjective* measure (e.g. satisfaction with own income). Subjective measures of individual wellbeing can reflect the individual's *evaluation* across multiple dimensions of life (e.g. satisfaction with life overall) or incorporate the individual's *preferences* across those dimensions. The normative and empirical implications of these different concepts of well-being are discussed in Decancq et al. (2015a).

This paper reports on an attempt to operationalise a preference-based measure of overall individual well-being based on people's stated preferences over stylised hypothetical scenarios. This measure is the so-called "equivalent income" (Fleurbaey and Schokkaert, 2012; Fleurbaey et al, 2013), or "equivalent *consumption*". To illustrate this concept with just two dimensions – "health" (h) and "material consumption" (c) – assume an individual with a combination of health and consumption (h,c), and a level of overall wellbeing that is associated with the combination. Let $h*$ denote full health, and imagine another combination ($h*$,c') that is equally good *to the individual* as the current combination. If we assume that "full health" is the same across everybody, then the level of consumption c' alone represents the level of wellbeing that is equivalent to the individual's current level of wellbeing. This c' is equivalent consumption, an interpersonally comparable uni-dimensional measure of wellbeing which collapses health and consumption based on each individual's own preferences.

If the individual is already in full health ($h = h*$), then $c' = c$. However, if not ($h < h*$), then assuming individual utility is increasing in health and consumption, $c' \leq c$ to compensate. The difference, $c - c'$, corresponds to the reduction in current consumption the individual is willing to forego in exchange for full health – thus, the "willingness to pay (WTP) to be in full health". The size of equivalent consumption is a function of: how healthy one is; how much one cares about one's health; and how much consumption one has. The lower one's consumption is, the lower is one's equivalent consumption, adjusting for: one's level of health (the less healthy, the lower the equivalent consumption); and one's preference between health and consumption (the more one cares about health relative to consumption, the lower the equivalent consumption - unless already in full health).

In stark contrast to the other approaches on subjective wellbeing, equivalent consumption is a measure of individual wellbeing that does not require full interpersonal comparison of satisfaction *levels* - it is the equivalent consumption levels that are interpersonally comparable, not the utility levels. It therefore avoids some of the well-known criticisms that have been raised against subjective welfarism, e.g. in the work of Amartya Sen (1985). More specifically, it does not suffer from the so-called "physical-condition neglect", i.e. the problem that individuals adapt their aspiration levels to their actual situation.

Equivalent consumption is a ratio-scale measure, expressed in money terms. It can therefore be introduced in any concave, i.e. inequality-averse, social welfare function. This suggests an alternative (non-welfarist) way to introduce distributional weights in economic evaluation (Fleurbaey et al. 2013). However, applying equivalent consumption raises a difficult informational challenge: we have to measure individual preferences. Different methods have been proposed in the literature. Decancq et al. (2015b) derive information about marginal rates of substitution from estimated happiness equations, Fleurbaey et al. (2013) implement a contingent valuation approach. This research is still at an experimental stage and, this paper reports on an informal pilot study that aimed to operationalise equivalent consumption as a measure of individual utility and of social welfare using choice experiments.

There are five objectives:

1. to examine if equivalent consumption can be operationalised through choice experiments at the individual and the aggregate levels;
2. to administer the experiment in an online survey;
3. to investigate qualitatively, how respondents interpret and approach the online survey;
4. to capture the effect of the mode of administration of the survey across online and face to face interview (although numbers are small); and
5. to explore whether the framework can be used to go beyond the original "individual" perspective and implement a societal perspective.

To address the **first objective**, we look at two methods to operationalise equivalent consumption. Continuing with the two-dimension example above, the *first method* is to ask respondents to give their WTP for full health (h^*). This is the method that was already applied in Fleurbaey et al. (2013). The *second method* is an indirect approach based on asking respondents to assess hypothetical scenarios combining different levels across dimensions of wellbeing that they might find themselves in; estimate the individual's utility function; and work out the combination (h^*,c') that is equivalent to the current situation (h,c) based on the coefficients of the utility function. Most studies that model individual utility functions pool data across individuals, and ignore individual level heterogeneity in preferences. At best, they may divide individuals into subgroups with similar preferences and thus only account for group level heterogeneity. In this paper, in addition to modelling pooled data, we attempt to model individual preferences at the level of each respondent. Estimating utility functions for each individual means allowing for inter-individual heterogeneity in preferences.

Regarding the **second objective**, if equivalent consumption can be operationalised through online surveys, this would open up revenues for cost effective primary data collection. However, the exercise may be too complex for surveys without an interviewer present. The paper reports on the results of administering the survey online, using a convenience sample of non-academic university staff.

Regarding the **third objective**, the online survey may generate results that are amenable to quantitative analysis, but this will not be sufficient to assess the validity of the results. We conduct a qualitative interview survey to investigate: how respondents approach the tasks; and whether they believed that wellbeing factors are commensurable and could be traded against each other, and in particular, against monetary consumption. The respondents are recruited from the same population as the online sample.

Regarding the **fourth objective**, the respondents to the qualitative interviews are asked to complete the online survey as part of the interview. This allows a comparison of the results by mode of administration. If the two results are substantially distinct from each other, there are at least two possibilities. One is that the survey is too complex for respondents to complete on their own. The other is that the presence of the interviewer in the interview setting makes people give "Sunday-best" responses. (However, it should be noted that sample size is small.)

Regarding the **fifth objective**, instead of presenting to respondents the hypothetical scenarios as life situations that they might find themselves in, one could also present the scenarios as illustrations of life situations of other people. Asking respondents to assess these scenarios using a societal perspective (Dolan et al, 2003; Tsuchiya and Watson, 2017) would result in a social welfare ranking, to which an appropriate social welfare function could be fitted. However, it should be noted that this is a diversion from the equivalent consumption approach. It is essential to that approach that social welfare is a function of individual equivalent consumption, which respects the trade-offs between own health and own income as represented by the individual preferences. But aggregation across individuals is a matter of justice and the parameter of inequality aversion cannot be derived directly from individual preferences. Indeed, one could question the idea that social welfare judgements should necessarily respect individual preferences over their own lives. Thus, it is interesting to see if

this social evaluation exercise yields trade-offs that are similar to the ones that follow from each individual's evaluation of her own life.

2. METHODS

2.1 The survey

In this paper we use the term "equivalent *consumption*" rather than "equivalent *income*", although the latter is used more often in the literature. Our terminology is because the reason for including a monetary dimension in the wellbeing measure is to represent the person's objective standards of living and access to material goods, rather than the standing of the individual in the labour market. It is not our intention to imply that the absolute and relative size of incomes have no effect on individual subjective wellbeing beyond the capacity to consume that they entail – our intention is to capture material standards of living in the form of consumption.

Furthermore, since equivalent consumption is a measure of *individual* wellbeing, the measure of consumption it builds on also needs to be for each individual, as opposed to household. However, since a large proportion of the population live in multi-person households, this involves a practical difficulty. Moreover, ideally, equivalent consumption should be operationalised to capture the level of individual *lifetime* wellbeing – but in this study, the duration component is not addressed explicitly.

The concept of equivalent consumption can be operationalised across two or more attributes of wellbeing (X) against consumption. We use three attributes from the health domain (physical function, pain, and depression), and close relationships. The level of equivalent consumption for each individual is obtained in two approaches. The *first approach* asks for the WTP to achieve X^*, the best combination of the non-monetary attributes. The questionnaire instructs respondents to imagine their problems in health and relationships went away, and asks for the percentage reduction in personal spending (viz. personal consumption) that would leave them as satisfied as they currently feel. Equivalent consumption for each respondent can be calculated by combining this with information on the respondent's current situation (X,c).

This is a variant of the approach used in Fleurbaey et al. (2013), where WTP for full health was asked in terms of an absolute amount of money. However, people may not be willing to trade between consumption and non-material dimensions of wellbeing. Furthermore, if the results are to be used to evaluate intermediate outcomes (e.g. best level of physical function and pain, but second-best level of depression and relationships), then it would be necessary to: either assume a linear utility function; or to estimate the utility function allowing for non-linearity. In order to model individual level preferences, the latter would require equivalent consumption data for a number of intermediate outcomes. An alternative is to pool the data for different respondents: it is then possible to parameterise differences between groups of individuals, but it is no longer possible to model preferences at the individual level.

The *second approach* to obtain the level of equivalent consumption for each individual is indirect and uses a variant of choice experiments using hypothetical scenarios to explore the implied levels of equivalent consumption for intermediate outcomes. Estimating the utility function across three dimensions at the individual respondent level through pairwise choice experiments would require a large number of choice tasks from each individual. However, if respondents are asked a small ranking task of, say, four hypothetical scenarios, A, B, C, and D ("$J = 4$" in the choice experiment jargon), then these data can be "exploded" into six implied pairwise choice observations (4 x 3 x ½ = 6). If respondents answer a series of six such ranking tasks ("$S = 6$"), that can be exploded into 36 pairwise choice tasks: much more than is asked in usual binary choice experiment surveys. This approach,

which we used, is called "sequential best-worst scaling" (Louvier, et al, 2008), but let us refer to it as "sequential ranking" (which we think is more self-explanatory).

Since the objective of the study is not to ascertain what the most important dimensions of wellbeing are, the attributes, the number of levels and the wording used in the choice exercise were determined by the research team (IA, ES and AT) based on judgement. The attributes and levels used are summarised in Table 1.

Table 1 The attributes and levels used in the sequential ranking tasks

Attribute	Levels
Physical functioning	I have no problems [0] / some problems [1] / extreme problems with physical functioning [2]
Pain	I have no pain [0] / some pain [1] / extreme pain [2]
Depression	I am not depressed [0] / moderately depressed [1] / extremely depressed [2]
Close relationships	I do not have relatives and friends with whom I can discuss intimate and personal matters [0] / I have… [1]
Personal spending	£6,000 p.a. / £20,000 p.a. / £41,000 p.a. [continuous]

The survey consisted of seven sections:

a) Survey information and consent;
b) Respondent background questions (age, sex, education, physical function, pain, depression, close relationships);
c) Personal spending for the past 12 months;
d) Global life satisfaction;
e) Direct equivalent consumption as percentage of current spending;
f) Sequential ranking tasks; and
g) Feedback questions.

For the questions in section (b) on physical function, pain, depression, and close relationships, the levels used in the sequential ranking tasks were used.

For personal spending for the past 12 months in section (c), respondents were first asked whether they would like to work in: weekly and annual amounts; or in monthly and annual amounts. They were then asked questions on take-home household income, household spending, number of people in the household, and personal spending. Regarding the three monetary questions, to avoid respondents having to give an exact figure (which may be imprecise), and to avoid the use of ranges (which is problematic to analyse), respondents were asked to select one option that came closest from a number of alternatives. Five alternatives ranging from £11,000 to £64,000 p.a. were given for take-home household income, based on the mean household disposable income figures by quintiles for 2012/13 (ONS, 2014a) and rounded. The decision to use five options is ultimately arbitrary: if too many options are given, respondents will spend too much time trying to select the appropriate option and/or the choice may be subject to error; if too few options are given, error may be reduced but precision will be sacrificed. For household spending, given that average savings ratios in the UK has averaged around 5-8% over the recent years (ONS, 2014b), it was decided to use the same options as the income question. For personal spending, the main consideration was the number of people in the household and how respondents might adjust for this. Theoretically, some proportion of household spending will be shared spending (e.g. utility bills, mortgage repayment), so personal spending on average should be higher than the household spending divided by household size. However, we also need to allow for some respondents to respond in such a manner. Therefore, the option at the top end was fixed at £65,000 p.a. for the single person with the largest spending, and the lowest end was set at £2,000 p.a. to allow for a respondent in the lowest household spending category in a five-person

household calculating personal spending as per capita spending. Across these two extremes, eight more options were allocated.

Life satisfaction in section (d) was asked as: "All things considered, how satisfied have you been the past 12 months with your life as a whole?" This was followed by a 10-point scale labelled "Not at all satisfied" at 1 and "Completely satisfied" at 10.

Direct equivalent consumption in section (e) was asked in two steps. The first step (WTP1), as a warm up, asked "Suppose your current health and relationship problems went away, but your spending was cut to half its current level, would you say you would be…" with three options: better off; about the same; and worse off. The second step (WTP2) starts with the same assumption and continued: "what would be your personal spending level that would make you feel as satisfied with your life as during the past 12 months?" There were 11 options ranging from zero WTP (same spending as now) to maximum WTP (zero spending) in 10% notches.

The sequential ranking exercise in section (f) started with an example, followed by the ranking tasks, each involving four hypothetical "lives" described in terms of physical functioning, pain, depression, close relationships, and personal spending. The health and relationship attributes used the same levels as the respondents' background questions. The personal spending attribute used the second, fifth and eighth options available in the respondents' personal spending question. The health and relationship attributes were treated as categorical, and the spending attribute as continuous. The data were modelled with interactions between each of the categorical attributes and the spending attribute, and spending squared. This implies a model with 16 parameters (eight for the main effects, plus seven for the interactions and one squared term). Attribute level balance would mean a survey with either six or 12 tasks. For a choice experiment involving sequential ranking of four scenarios at a time ($J = 4$) with 16 parameters to estimate ($K = 16$), given a sufficiently large sample size, the minimum number of tasks the design should have is $S = K / (J - 1)$, which in this case is 5.33. On the one hand, estimating an individual model means sample size will be one, and six tasks are likely to be insufficient for the regression model to converge. But on the other hand, the authors (IA, ES and AT) each tried out 12 sequential ranking tasks and agreed the volume is very challenging. In the absence of further information, the experimental design software Ngene (Choice Metrics, 2012) was used to draw up two designs based on D-efficiency: one with six and the other with 12 ranking tasks (both with $J = 4$). A screenshot of the explanation of the task, and of an example task are shown in the Appendix.

There were two versions of the questionnaire, and the only difference was in the perspective of the sequential ranking tasks (Tsuchiya and Watson, 2017). *Version IU* was for individual utility: respondents were told that lives A to D were imaginary lives that they may find themselves in, and were asked to rank them from the best to the worst. *Version SW* was for social welfare: the questionnaire stated that "one important role of government is the provision of welfare benefits and public aid […]. If government is to operate in a fair manner, people who are better off should pay more taxes, and people who are worse off should be entitled to more benefits." Respondents were then told that lives A to D described the lives of four different people, and were asked to rank them from the best off to the worst off, with the above role of government in mind. Neither version explicitly referred to duration of these lives or what happens afterwards.

The feedback questions in section (g) had two parts. The first part presented 11 statements and asked the respondent to indicate all that applied. The statements included for example: "Too many tasks", "Got tired half way through", "Not sure about my answers". Version SW had two additional statements: "Imagined the lives happening to myself" and "Difficult to judge other people's lives". The order in which the statements appeared was randomised across respondents. The second part was an optional text box asking for free text comments.

The online survey was designed as a "Google form". The two versions, IU and SW, were set up as separate forms, and a separate webpage was set up to randomise respondents between the two versions. Screenshots of the whole questionnaire is available from AT.

2.2 The face to face interviews

The face to face interviews consisted of respondents answering section (b) onwards of the survey on-line themselves, but in the presence of CS. From section (c) onwards, after they answered the tasks, respondents were invited to comment on what they thought of the questions and how they answered them. After section (g), the respondents were asked whether they had a timeframe in mind when they approached the choice tasks. Respondents were also informed of the two versions (IU and SW), and asked whether their approach and answers would have differed had they been given the other version. The interviews were audio-recorded for transcription.

2.3 KU Leuven pre-pilot

Prior to the actual survey, a small scale paper-and-pencil pre-pilot of the survey was carried out at KU Leuven in May 2014, to examine how many sequential ranking tasks respondents could cope with, and whether the data can be modelled at the individual level. This subsection reports how this pre-pilot resulted in the decision to use six sequential ranking tasks in the actual survey.

The individual utility version of the questionnaire (in English) was conducted in an undergraduate Economics class taught by ES. Given the very small sample size and the impossibility of guaranteeing anonymity, the respondents were asked to read through and give due consideration to, but not fill in their answers to, the factual questions (self-reported health, relationships, spending, life satisfaction, WTP1, WTP2), and then actually answer the 12 sequential ranking tasks. Ten students attended the class, one student started but did not complete the questionnaire, resulting in responses from nine students.

The individual level main effects model converged for six out of nine respondents. The models were significant overall, but few coefficients were significant. With one exception, the significant coefficients were for the worse levels and had the expected sign. The individual level interaction model converged for five out of nine respondents, but none of the coefficients were significant. Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC), both of which assess the size of the residual against the number of explanatory variables (the smaller the better), indicated the model was not improved by the addition of interaction terms. When the data were pooled across all nine respondents, the model performance improved substantially, resulting in seven (out of eight) significant coefficients, all with the expected sign. Adding the interaction terms to the pooled model results in four (out of 15) significant coefficients, while AIC and BIC indicated that the addition did not improve the model.

While the pre-pilot indicated that 12 sequential ranking tasks can be completed by Economics students in the presence of a facilitator, it was not clear that the 12 tasks could also be completed by a less technical sample, following extensive questions on spending, and in an online environment. It was therefore decided that the design with six sequential ranking tasks is used in the online survey. This should enable more respondents to complete the survey, but may also mean the individual level model converges only for fewer respondents.

2.4 The two samples

Respondents to the main online survey (Survey 1) were recruited using the e-mail announcement system at the University of Sheffield. A message inviting volunteers to access the online survey was sent to e-mail addresses of staff with jobs in administration (clerical, secretarial, managerial, and

professional) and facilities (catering, cleaning, porters, security, etc). No incentives were offered. Respondents were randomised to Version IU and Version SW when they accessed the survey on-line.

The interview sample (Survey 2) was recruited using the same method. The message was sent to the e-mail addresses of the same categories of staff, most of whom would also have received the first invitation message. The e-mail outlined the interview and invited only those with no postgraduate qualifications (to achieve a more homogeneous sample) - those who wished to take part were asked to contact CS to arrange for an interview. Participants to Survey 2 were each offered a £10 high street voucher.

2.5 Quantitative analysis

Data from sections (b) (c) (d) (e) and (g) are reported as descriptive statistics. The overall sample is broken down by survey (1. online; 2. Interview), and by version (IU. individual utility; SW. social welfare). Furthermore, within the online sample, those with no post-graduate qualifications form a subsample that is directly comparable to the interview sample in terms of academic qualifications.

Data from sections (c) and (e) are used to calculate direct equivalent consumption in monetary terms for each respondent by multiplying their personal spending by (1- WTP2). The quality of the direct WTP questions is looked at in two ways. *First*, regarding WTP1, those respondents with problems in health or relationships may be better off, about the same, or worse off at (X^*, $0.5c$); but those with no health or relationship problems should reply that they would be worse off. Therefore, respondents' self-reported health and relationship status (X) are cross-tabulated against their response to WTP1. Note that there are three possibilities. (i) We observe $X < X^*$ because the respondent has (and reports) at least one problem in health or relationships. (ii) We observe $X = X^*$ because the respondent has (and reports) no problems in health or relationships. (iii) We observe $X = X^*$ because the respondent has a medical condition but reports no problems because they have no physical symptoms, pain or depression. Given the data, we cannot distinguish between the second and third possibilities.

Second, those who responded they would be better off / about the same / worse off in WTP1 should be indifferent to loose (or willing to give up) more than / around / less than 50% of their consumption, respectively. Therefore, respondents' responses to WTP1 are cross-tabulated against their response to WTP2, separately by their health and relationship status.

The modelling of the sequential ranking tasks assumes that the probability of a given scenario being ranked first (most preferred) is a function of the utility associated with the scenario, which is made up of observable components (i.e. the levels of the five attributes) and an unobserved component. Thus, data from section (f) are analysed using rank ordered logit regressions in two ways. *First*, N individual level models are estimated. The dependent variable has J x S observations from each respondent, which are grouped by ranking task (so, S groups). *Second*, aggregate level models are estimated using pooled data. The J x S observations from N respondents are stacked to form the dependent variable, and analysed grouped by respondent-task (so N x S groups). Six pooled models are run using different samples: the whole sample from Survey 1, the no-PG qualification subsample from Survey 1, and Survey 2; each broken down by the two versions (IU, personal; and SW, social perspectives).

Due to the way Google Form is organised, respondents could choose only one life (A, B, C, or D) at a given ranking slot (e.g. the most preferred), but they could still give two (or more) ranks to one life and not rank another life (e.g. life A is the most preferred *and* the second most preferred). When a respondent does this, it is not possible to discern whether it is a simple error or a misunderstanding of the exercise. Individual level models are run regardless. Pooled models are run without data from tasks involving such inconsistent data.

Furthermore, the main effects results from section (f) are combined with the results from section (c) to calculate predicted equivalent consumption at the individual level. The econometric modelling of the sequential ranking data assumes a main effects additive objective function over hypothetical lives (L = A, B, C, D) of $u_L = \beta_1 X_L + \beta_2 c_L$. The definition of equivalent consumption implies: $\beta_1 X_r + \beta_2 c_r = \beta_1 X^* + \beta_2 c'$; where subscript r indicates the individual's real current situation. Thus, where spending is measured in linear units, individual equivalent consumption (c') given current situation (X_r) can be predicted by: $c' = \frac{\widehat{\beta_1}}{\widehat{\beta_2}} X_r + c_r$. (Note that X^* = 0.) However, this assumes that the marginal rates of substitution between consumption and X are constant, which may not hold. To accommodate non-linearity in consumption, econometric modelling uses the log of spending, in which case equivalent consumption becomes:

$$c' = exp\left[\frac{\widehat{\beta_1}}{\widehat{\beta_2}} X_r + \ln(c_r)\right].$$

This formula is used to calculate the indirect measure of equivalent consumption for each individual respondent. An implication is that equivalent consumption will always be positive.

To compare the effect of perspective, the pooled regression results of the two versions (IU and SW) from survey 1 are visually inspected. Furthermore, the whole sample across the two versions is modelled together. The main effects variables are interacted with a version dummy to explore the effect of version on the results. However, note that the choice experiment design used in the survey did not build in the possibility of such an analysis and therefore the analysis is only exploratory.

2.6 Qualitative analysis

The audio-recorded interviews were transcribed verbatim professionally. These were subjected to thematic content analysis, using the Framework method (Ritchie, Lewis, 2005), to identify key themes. CS and KS independently coded three interview scripts, discussing their coding after the coding of each script. After agreement on the coding framework, CS coded the remaining scripts and discussed the results of this with KS and AT.

3. RESULTS

3.1 The respondents

For Survey 1, the invitation e-mail to non-academic non-researcher members of staff at the University of Sheffield was sent to 3,371 e-mail addresses in September 2014. The survey was open for one week with no reminders, and had a total of 96 accesses, of which 74 were on the first day. In total, 52 respondents completed it, indicating a 46% dropout rate across the two versions (breakdown by version is not available).

For Survey 2, the invitation message was sent to 2,903 e-mail addresses using the same channels, in early July 2015. A total of 19 individuals contacted CS, and 17 interviews were conducted by the end of July 2015. Of the remaining two, one person cancelled the interview and one was left out due to a professional relationship with the interviewer. We do not know if any respondent completed both Survey 1 and Survey 2 (none of the Survey 2 respondents volunteered such information). The interviews alternated between Version IU and Version SW. Of the 17 respondents, 16 completed the online survey within the interview in CS's presence, but on their own. Interview respondent no.6 only completed three of the six sequential ranking tasks on-line due to lack of time (but completed the rest of the interview survey).

The background characteristics and direct WTP from the two surveys are summarised in Table 2.

Table 2 Descriptive statistics

Variable	Options	Survey 1 (on-line)				Survey 2 (interview)	
		All		Non-PG[(1)]			
		IU n=22	SW n=30	IU n=12	SW n=18	IU n=8	SW n=9
Age	20 or less	**8**	**12**	**6**	**9**	0	0
	21 to 35	7	6	3	2	2	**4**
	36 to 50	7	**12**	3	7	**6**	**4**
	51 to 65	0	0	0	0	0	1
	66 or above	0	0	0	0	0	0
	Prefer not to say	0	0	0	0	0	0
Sex	Male	9	4	6	2	2	1
	Female	**12**	**26**	6	**6**	6	8
	Prefer not to say	1	0	0	0	0	0
Education	No qualifications	0	0	0	0	0	0
	CSE / GCSE / O levels	1	4	1	4	1	0
	A levels, higher education below degree, or equivalent	2	2	2	2	**3**	**4**
	Degree or equivalent	9	**12**	9	12	3	4
	Postgraduate qualification	**10**	**12**	0	0	0	1
	Other	0	0	0	0	1	0
	Prefer not to say	0	0	0	0	0	0
Physical functioning	No problems [level 0]	**13**	**21**	**7**	**13**	**7**	**7**
	Some problems [1]	9	9	5	5	1	2
	Extreme problems [2]	0	0	0	0	0	0
Pain	No pain [0]	**12**	**15**	**7**	**9**	**5**	**5**
	Some pain [1]	10	**15**	5	**9**	3	4
	Extreme pain [2]	0	0	0	0	0	0
Depression	Not depressed [0]	**17**	**20**	**9**	**11**	**8**	**8**
	Moderately depressed [1]	5	10	3	7	0	1
	Extremely depressed [2]	0	0	0	0	0	0
Close relations	I have relatives and friends [0]	**22**	**29**	**12**	**17**	**8**	**9**
	I do not have relatives and friends [1]	0	1	0	1	0	0
Profile	Reports no problems ($X = X^*$)	9	**15**	**6**	**9**	**5**	4
	At least one problem ($X < X^*$)	**13**	15	6	**9**	3	**5**
Household take-home income	£11,000	1	1	1	0	0	1
	£19,000	4	8	2	**7**	**3**	1
	£25,000	5	3	**4**	2	1	1
	£35,000	**6**	**9**	3	6	**3**	**5**
	£64,000	**6**	**9**	2	3	1	1
	Mean household take-home income [(2)]	£36,565	£37,633	£31,833	£32,500	£31,375	£32,667
Household spending	£11,000	3	4	3	1	1	1
	£19,000	4	7	2	**7**	2	2
	£25,000	4	6	3	5	1	**6**
	£35,000	**8**	**8**	**4**	3	**3**	0
	£64,000	3	5	0	2	1	0
	Mean household spending [(2)]	£30,955	£30,900	£23,833	£27,889	£30,375	£22,111
Household size	1	5	6	4	3	**2**	1
	2	**11**	**14**	**7**	**11**	1	**7**
	3	1	2	0	0	**2**	1
	4	3	6	0	3	**2**	0
	5	1	1	1	1	0	0

	6	0	0	0	0	0	0
	7	0	0	0	0	0	0
	8+	0	0	0	0	0	0
	Mean household size [2] [3]	2.2	2.4	2.1	2.3	2.6	2.0
Personal spending	£2,000	2	2	2	2	1	1
	£6,000	4	4	**3**	0	2	1
	£10,000	3	6	1	4	**3**	**4**
	£15,000	**5**	**10**	2	**8**	2	1
	£20,000	3	4	2	2	0	2
	£26,000	3	2	2	2	0	0
	£32,000	1	1	0	0	0	0
	£41,000	0	1	0	0	0	0
	£51,000	0	0	0	0	0	0
	£65,000	1	0	0	0	0	0
	Mean personal spending [2]	£16,727	£14,766	£12,833	£14,222	£9,250	£11,444
	Mean per capita household spending [3] [4]	£16,576	£14,948	£14,063	£13,389	£1,4857	£11,204
Life satisfaction	1	0	1	0	1	0	0
	2	1	0	1	0	0	0
	3	0	2	0	2	0	0
	4	2	2	0	2	0	0
	5	1	1	0	1	0	0
	6	2	2	2	2	1	0
	7	3	4	1	**4**	1	3
	8	**8**	**9**	**5**	2	**4**	**5**
	9	5	6	3	3	2	1
	10	0	3	0	1	0	0
	Mean life satisfactions	7.1	7.2	7.3	6.3	7.9	7.8
WTP1 [5]	Better off than now	4	7	2	5	1	2
	About the same as now	8	6	**6**	3	2	2
	Worse off than now	**10**	**17**	4	**10**	**5**	**5**
WTP2 [6]	0% of current spending (not willing to pay anything)	**10**	**19**	**8**	**11**	**3**	**5**
	10% less spending	1	2	0	2	0	1
	20%	3	4	1	2	2	3
	30%	3	1	0	1	1	0
	40%	2	2	1	1	0	0
	50%	2	1	1	1	1	0
	60%	0	0	0	0	1	0
	70%	0	1	0	0	0	0
	80%	0	0	0	0	0	0
	90%	0	0	0	0	0	0
	100% (willing to give up all spending) [2]	1	0	1	0	0	0
	Mean WTP percentage of current spending	20%	11%	18%	10%	22%	8%
ecdir [7]	Mean direct equivalent consumption	£12,677	£13,456	£8,817	£13,139	£7,163	£10,533
	SD	£13,145	£8,493	£6,276	£6,810	£4,486	£5,586

Notes

Modal response category for each question by subsample is in bold.

(1) The "non-PG" group is a subsample within "all".

(2) Mean of the options selected by the respondents.

(3) Two respondents in survey 1 (one from each variant) did not report household size.

(4) Per capital household spending calculated by dividing each respondent's reported household spending with their reported household size.

(5) "Suppose your current health and relationship problems went away, but your spending was cut to half its current level, would you say you would be …?"

(6) "Suppose your current health and relationship problems went away, what would be your personal spending level that would make you feel as satisfied with your life as during the past 12 months?"

(7) Direct Equivalent Consumption: calculated for each individual from own personal spending and percentage WTP to achieve X^*

3.2 Direct equivalent consumption through willingness to pay

Direct equivalent consumption is calculated for each individual using their reported level of personal spending. In Survey 1, there is a high correlation (ρ = 0.82) between per capita spending (household spending divided by household size) and personal spending, suggesting that at least some respondents may have equated personal spending with per capita spending. The correlation is much lower in Survey 2 (ρ = 0.013).

Table 2 above illustrates that five of the six columns have similar distributions in WTP, with around half the respondents answering worse off than now for WTP1. The modal willingness to forego current consumption in WTP2 is 0% across all six columns. For those who report a WTP2 of 0%, equivalent consumption is identical to their personal spending. Any discrepancy in WTP across the two versions (IU vs. SW) is not attributable to the version/perspective, since the two versions are identical up to this stage. The last row gives the mean direct equivalent consumption calculated for each respondent from their WTP2, i.e. based on their personal spending and the proportion of their personal spending they are willing to give up in order to achieve X^*.

Figure 1 presents scatter plots of direct equivalent consumption against personal spending, by survey. Note that the two panels are scaled the same, but all observations from Survey 2 have personal spending ≤£20,000.

Figure 1: Scatter plot between direct equivalent consumption and personal spending

(a) Survey 1 (n=52) (b) Survey 2 (n=16)
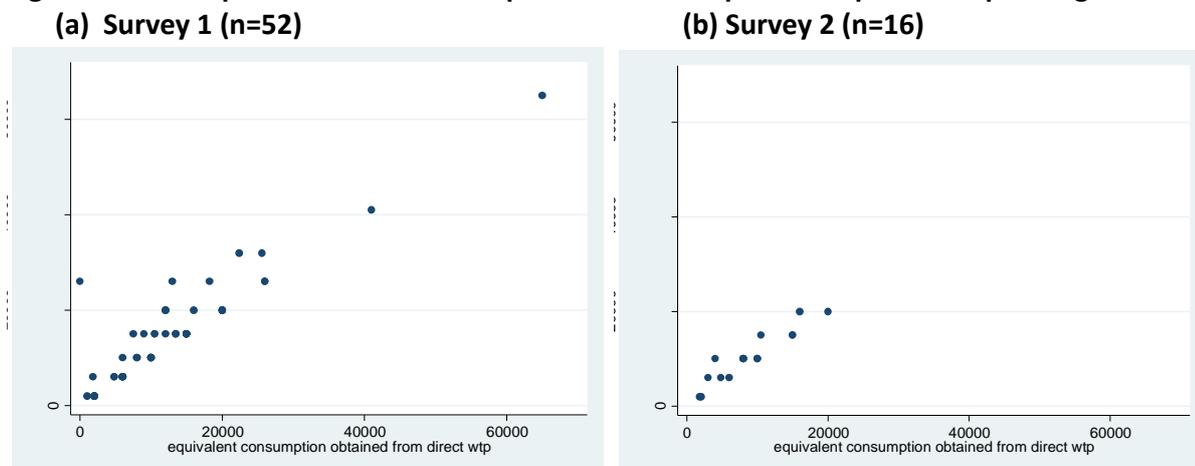
Regarding the quality of these WTP data, Table 3(a) cross-tabulates self-reported health and relationship status (X) against WTP1. Since the two versions of the survey are identical up to the WTP questions, the two versions are pooled within each survey to analyse these. Assuming those who do not report any health or relationship problems ($X = X^*$) actually have no problems, it would be inconsistent for such respondents to reply "about the same / better off than now" for WTP1, because it would imply a strictly negative marginal utility of personal spending (however, as was noted above, they may have a asymptomatic medical problem, which they are willing to pay a positive amount to be cured of). The shaded cell indicates such a *potentially* inconsistent combination, with ten respondents (19%) from Survey 1 and six (35%) from Survey 2. The proportions are not statistically significantly different across the surveys (z-test; p = 0.17).

Table 3 (a): Self-reported problems in health and relationships against WTP1 [1]

Any problems? / WTP1	Survey 1			Survey 2		
	$X < X^*$	$X = X^{*\,(2)}$	Total	$X < X^*$	$X = X^{*\,(2)}$	Total
Better off than now: $(X^*, 0.5c) \succ (X, c)$	8	3	11	1	2	3
About the same as now: $(X^*, 0.5c) \sim (X, c)$	7	7	14	0	4	4
Worse off than now: $(X^*, 0.5c) \prec (X, c)$	**13**	**14**	27	**7**	**3**	10
Total	28	24	52	8	9	17

Notes

(1) Modal cell by X status is in bold. Highlighted cells imply *potential* logical inconsistency.

(2) Those reporting no problems in physical functioning, no pain, and no depression.

Table 3(b) cross-tabulates the responses to WTP1 against responses to WTP2, for those with problems ($X < X^*$). If, for example, a respondent with $X < X^*$ indicates in WTP1 that $(X^*, 0.5c) \succ (X, c)$, then since $(X^*, c') \sim (X, c)$ by definition, therefore in WTP2 they must indicate $c' < 0.5c$. In other words, replying "better off than now" in WTP1 should be associated with a willingness to give up more than 50% of current spending in WTP2. Or, the other way round, and $(X^*, 0.5c) \prec (X, c)$ implies $c' > 0.5c$.

In Survey 1, of the 28 respondents with $X < X^*$, 13 (46%) fall in a shaded cell, indicating internally inconsistent combinations; eight (62% of those in the shaded cells) are in the 0% row: i.e. not willing to give up any consumption in exchange for improved health and close relationships. In Survey 2, of the eight respondents with $X < X^*$, one (13%) falls in a shaded cell, in the 0% row. The proportions falling in a shaded cell are statistically significantly higher at 10% for Survey 1 than for Survey 2 ($p = 0.08$).

Table 3(b): WTP1 against WTP2 for those $X < X^*$

WTP1 / WTP2[1]	Survey 1				Survey 2			
	Better off	About the same	Worse off	Total	Better off	About the same	Worse off	Total
0%	**4**	**4**	5	13	**1**	0	3	4
10% - 40%	2	2	**8**	12	0	0	**4**	4
50%	1	1	0	2	0	0	0	0
60% -100%	1	0	0	1	0	0	0	0
Total	8	7	13	28	1	0	7	8

Notes

Modal cell by WTP1 response category is in bold.

Highlighted cells imply logical inconsistency.

(1) Proportion of to current personal spending respondent is willing to give up to achieve X^*

Table 3(c) does the same for those who report no problems ($X = X^*$). Here, if a respondent actually had no problems in health or relationships, then we would not expect the respondent to be willing to give up any personal spending in WTP2. Of the 24 and nine respondents who report $X = X^*$ across the two surveys, eight (33%) and five (56%), respectively, are in the shaded cells, indicating a positive WTP. The proportions are not statistically significantly different across the two surveys ($p = 0.42$).

Table 3(c): WTP1 against WTP2 for those $X = X^*$

	Survey 1	Survey 2

WTP1 / WTP2[1]	Better off	About the same	Worse off	Total	Better off	About the same	Worse off	Total
0%	0	**6**	**10**	16	0	**3**	1	4
10% - 40%	**2**	1	3	6	**1**	0	**2**	3
50%	1	0	0	1	0	1	0	1
60% -100%	0	0	1	1	**1**	0	0	1
Total	3	7	14	24	2	4	3	9

Notes

Modal cell by WTP1 response category is in bold.

Highlighted cells imply *potential* logical inconsistency.

(1) Proportion of to current personal spending respondent is willing to give up to achieve $X*$

3.3 Modelling the sequential ranking data

Table 4 summarises, by survey and by version, the incidence of inconsistent tasks, where a respondent gives two different rankings to at least one life and as a result does not rank all four lives in a given ranking task. While 12 respondents (23%) in Survey 1 had at least one inconsistent task, none of the respondents in Survey 2 had any inconsistencies. The difference in proportions across the surveys is significant (p = 0.03).

Table 4: number of inconsistent tasks per respondent

	Survey 1				Survey 2	
	All		Non-PG		IU (n=8)	SW (n=9)
	IU (n=22)	SW (n=30)	IU (n=12)	SW (n=18)		
0	18	22	8	12	7	9
1	2	5	2	4	0	0
2	0	2	0	1	0	0
3	1	0	1	0	0	0
4	0	0	0	0	0	0
5	0	0	0	0	0	0
6	1	1	1	1	0	0
Aborted	0	0	0	0	1[1]	0

Notes

(1) One respondent only answered three choice tasks due to insufficient time.

Table 5 summarises the individual level regression results. The columns indicated IU are for the personal perspective, and columns SW for the societal perspective. The row headings name the variables with the expected sign of the coefficient in parentheses. The numbers in the cells indicate the number of coefficients with p < 0.1. The two bottom rows indicate the number of models: the numbers run (the number of respondents) and the numbers that converged, reporting meaningful coefficients. For example, the first column illustrates that, of the 22 individual level models run for Version IU from online Survey 1, five of them converged, and across these, there were two models with a negative coefficient for level-2 *pain*, three models with a negative coefficient for level-2 *depression*, and one model with a positive *spending* coefficient, each with p < 0.1. Across the two surveys, most of the individual level models did not converge. This confirms that it is very ambitious to estimate individual preferences (even with an interviewer present). Where the model did converge, most of the significant (at p < 0.1) coefficients were for level-2 *pain*, *relationships*, and *depression,* and *personal spending*. Across all 152 coefficients in the 19 converged models there are no significant coefficients with the "wrong" sign. The only anomaly is that one respondent in Survey 1, Version SW

(respondent no.1) had a significant coefficient for level-1 but a non-significant coefficient for level-2 *physical functioning*. (Full results available on request.)

Table 5: summary of individual level regressions: numbers of coefficients with p < 0.1

Survey	Survey 1				Survey 2	
Versions	**All**		**Non-PG**		**IU**	**SW**
	IU	**SW**	**IU**	**SW**		
Physical functioning some (-)	0	1	0	0	0	0
Physical functioning extreme (-)	0	0	0	0	1	0
Pain some (-)	0	0	0	0	0	0
Pain extreme (-)	2	4	1	0	1	1
Depression some (-)	0	0	0	0	1	0
Depression extreme (-)	3	3	2	0	1	0
Relationship none (-)	1	2	1	0	1	1
Personal spending (+)	1	3	0	3	1	0
The number of models run	**22**	**30**	**12**	**18**	**7**	**9**
The number of models converged	**5**	**11**	**3**	**4**	**2**	**1**

Notes:
A: Version IU, asked from the personal perspective
B: Version SW, asked from the societal perspective

Table 6 reports the pooled regression results (all of which converged). All of the significant coefficients have the expected sign, and where both the level-2 and level-3 coefficients are significant they have the expected ordering. These imply minimum face validity of the results. With one exception (Survey 1, Non-PG, Version IU, personal spending) the significance of coefficients is common across the six models. With some exceptions, most significant coefficients have p < 0.01. Comparing the results for those without postgraduate qualifications across the two surveys, the Version IU coefficients for level-1 *depression* and *personal spending* are not significant at 5% for Survey 1, while significant for Survey 2; on the other hand, the Version SW coefficients are similar. However, also note that the Version IU models have smaller sample size. Panel (b) presents the rank ordering of the absolute value of the coefficients within each model. All models agree that: level-2 *depression* is the most important; and level 1 in *physical functioning* and *pain* have a statistically non-significant effects. *Personal spending*, and level-1 *depression* are also consistently less important. The relative importance of level 2 in *physical functioning* and *pain* and *relationships* vary.

Table 6(a): pooled regression models

	Survey 1				Survey 2	
	All		**Non-PG**			
	IU	**SW**	**IU**	**SW**	**IU**	**SW**
Physical functioning some (-)	0.129	0.116	0.211	0.226	-0.189	-0.137
Physical functioning extreme (-)	-1.627***	-0.749***	-1.982***	-0.904***	-1.889***	-1.345***
Pain some (-)	0.293	-0.01	0.132	0.088	0.188	0.231
Pain extreme (-)	-1.727***	-1.297***	-2.547***	-1.224***	-0.796***	-0.732***
Depression some (-)	-0.699***	-0.683***	-0.659*	-0.645***	-0.737**	-0.614**
Depression extreme (-)	-2.723***	-2.029***	-2.987***	-1.964***	-2.345***	-2.052***
Relationship none (-)	-1.135***	-1.250***	-1.444***	-1.278***	-1.474***	-1.234***
Personal spending (+)	0.358***	0.586***	0.289	0.531***	0.574***	0.576***
N of observations	484	660	244	384	180	216
ll	-226.428	-345.404	-97.27	-200.294	-87.338	-113.919

r2_p	0.411	0.341	0.498	0.343	0.389	0.336
df_m	8	8	8	8	8	8
chi2	316.234	357.949	193.183	209.598	111.349	115.392
aic	468.855	706.809	210.54	416.588	190.676	243.838
bic	502.312	742.747	238.517	448.193	216.22	270.84

legend: * p<.1; ** p<.05; *** p<.01; "ll" Log likelihood; "r2_p"_pseudo R^2; "df_m" model degrees of freedom; "chi2" χ^2; "aic" Akaike Information Criterion; "bic" Bayesian Information Criterion

Table 6(b): Ranking of absolute size of coefficients within each model

	Survey 1				Survey 2	
	All		**Non-PG**			
	IU	**SW**	**IU**	**SW**	**IU**	**SW**
Physical functioning some (-)	ns	ns	ns	ns	ns	ns
Physical functioning extreme (-)	3	3	3	4	2	2
Pain some (-)	ns	ns	ns	ns	ns	ns
Pain extreme (-)	2	2	2	3	4	4
Depression some (-)	5	5	5	5	5	5
Depression extreme (-)	1	1	1	1	1	1
Relationship none (-)	4	4	4	2	3	3
Personal spending (+)	6	6	ns	6	6	6
N of observations	484	660	244	384	180	216

3.4 Predicted equivalent consumption through sequential ranking

There are three major difficulties regarding the calculation of predicted equivalent consumption for each individual based on the regression coefficients of their own individual level models of the sequential ranking task.

1. Around half of respondents (47% of Survey 1; and 53% of Survey 2) report $X = X^*$, but this does not guarantee no health/relationship problems (and thus $c' = c_r$), because they may have other health/relationship problems not captured by the questions.
2. Most of the individual level models (69% of Survey 1; and 81% of Survey 2) did not converge.
3. The individual level models that did converge had few significant coefficients. Of the small set of respondents with $X < X^*$ *and* whose individual level model converged, none of the problems they had corresponded to a significant $\widehat{\beta_1}$ coefficient in the individual level regression model for the respondent (i.e. nobody with *some pain* had the level-1 *pain* coefficient significant; nobody with a significant level-1 *depression* coefficient had *some depression*; etc).

Instead, equivalent consumption was calculated indirectly for each individual using the *pooled* regression coefficients. Leaving aside the fundamental point, that equivalent consumption is a measure of wellbeing that incorporates *individual-specific* preferences, this approach still has practical difficulties (the first one is the same as above):

1. Around half of respondents report no health/relationship problems, but this does not guarantee $c' = c_r$.

2. The relevant (level-1) parameters with p < 0.1 are for *some depression* and *no relationships* – only 16 respondents (15 from Survey 1; and one from Survey 2) had problems in at least one of these dimensions.

Table 7 gives summary statistics by Survey and model. For example, the first row is for indirect equivalent consumptions calculated using the Version IU pooled coefficients from Survey 1 of those 15 respondents in Survey 1 with level-1 problems in *depression* and/or *relationships*. Their mean indirect equivalent consumption is £2,107.

Table 7: Indirect (and direct) equivalent consumption by Survey

		n	Mean (£)	SD (£)	Min. (£)	Max. (£)
Survey 1 All	Indirect IU		2,107	1,033	12	3,683
	Indirect SW	15	4,633	2,264	74	8,094
	Direct		12,787	6,877	1,000	26,000
		52	13,127	10,598	0	65,000
Survey 1 non-PG	Indirect IU		2,154	1,146	12	3,683
	Indirect SW	10	4,739	2,509	74	8,094
	Direct		13,050	7,228	1,000	26,000
		30	11,410	6,838	0	26,000
Survey 2	Indirect IU		2,767	–	–	–
	Indirect SW	1	3,441	–	–	–
	Direct		10,000	–	–	–
		17	8,947	5,236	1,800	20,000

Figure 2 is a scatter plot showing the relationship between the indirect and direct equivalent consumption figures, using the pooled coefficients. Those 15 respondents in Survey 1 with level-1 problems in *depression* and/or *relationships* are used to illustrate. The two versions are shown in separate panels. Indirect equivalent consumption is larger using the Version SW coefficients, because the spending coefficient, $\widehat{\beta_2}$, in the social welfare model is relatively larger than in the individual utility model (0.358 for Survey 1 IU, 0.586 for Survey 1 SW; Table 6(a)). As can be inferred from the two panels, indirect equivalent consumption calculated using the two sets of parameters are highly correlated (ρ = 0.999989) – this is because equivalent consumption is driven by actual consumption (which is the same across the two models), and because the $\widehat{\beta_1}$ coefficients are similar across the IU and SW models, so that most of the difference between the panels comes from $\widehat{\beta_2}$.

Figure 2: Scatter plot between direct and indirect equivalent consumption using respondents from Survey 1 with problems in depression and/or relationships

(a): Using Version IU parameters (n=15)   **(b) Using Version SW parameters (n=15)**
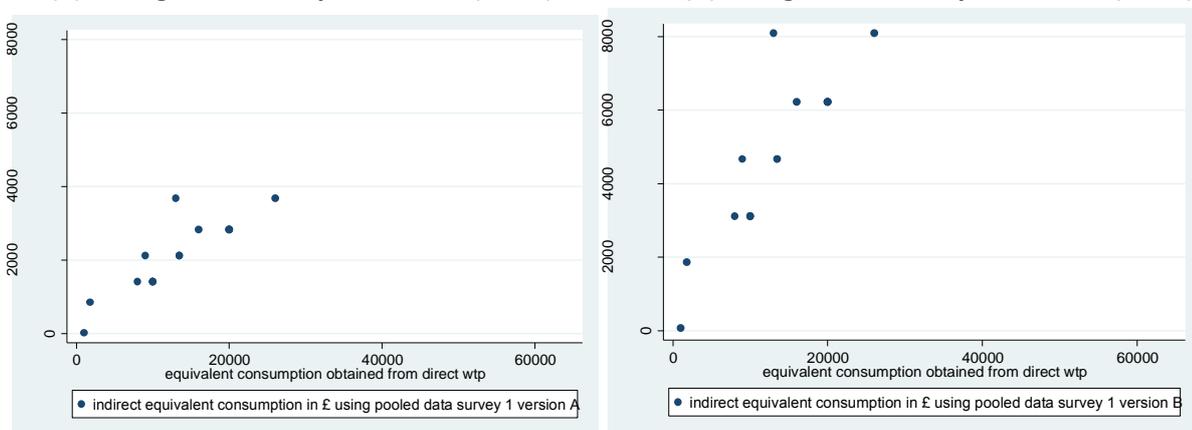
Table 8 reports equivalent consumption under different hypothetical states ($X < X^*$) combined with personal spending levels of either £6,000 or £41,000, assuming a person with the preferences represented by the six sets of pooled regression coefficients. The first row indicates that having *extreme problems in physical functioning* (but no other problems) and personal spending of £6,000 is equivalent to not having any problems and a personal spending ranging from £6 (Survey 1, Version IU, non-PG) to over £1,600 (Survey 1, Version SW, All). Across all models, having *extreme depression* (but no other problems) results in lower equivalent consumption than having any other (single) problem. Being in the worst state: *extreme problems in physical functioning*, *extreme pain*, *extreme depression*, and *no close relationships*, with either level of personal spending (£6,000 or £41,000) is equivalent to living in the best state (X^*) with hardly any personal spending (NB. the value is constrained to be positive). In other words, the associated willingness to pay amounts are, in effect, as large as the personal spending levels. Across the two surveys, Version IU consistently has smaller values, because the estimated coefficient of personal spending is smaller.

Table 8: Hypothetical indirect equivalent consumption for different X and personal spending

Hypothetical states		Survey 1				Survey 2	
		All		Non-PG		IU	SW
X	Personal spending	IU	SW	IU	SW		
Only physical functioning extreme	£6,000	£64	£1,668	£6	£1,093	£223	£580
Only pain extreme		£48	£655	<£1	£598	£1,500	£1,682
Only depression extreme		£3	£188	<£1	£148	£101	£170
Only relationship none		£252	£709	£40	£540	£460	£704
All of the above		<£1	<£1	<£1	<£1	<£1	<£1
	£41,000	<£1	£5	<£1	£2	<£1	£4

3.5 Tick box feedback

Table 9 summarises the tick box feedback. The proportions of negative statements within all statements are statistically significantly higher in Survey 1 than in Survey 2 ($p < 0.001$).

Table 9: Feedback from respondents

	Survey 1		Survey 2	
	IU (n=22)	SW (n=30)	IU (n=8)	SW (n=9)
Interesting exercise	11	16	8	8
Confident about my answers	8	3	3	2
I could answer 5 or 6 more of these tasks	2	0	4	3
Layout is clear	12	8	6	8
The task being asked is clear	12	10	7	7
Imagined the lives happening to myself [B]	-	5	-	8
Difficult to judge other people's lives [B][N]	-	17	-	6
Difficult to distinguish between the "lives" [N]	1	10	2	1
Got tired half way through [N]	4	8	0	0
Too many tasks [N]	1	5	0	0
I would not want to answer any more of these tasks [N]	11	15	0	2
Boring [N]	1	2	0	0
Not sure about my answers [N]	4	10	2	3
TOTAL[1] (average per respondent)	**67** (3.0)	**109** (3.6)	**32** (4.0)	**48** (5.3)
Total negative statements (percentage [2])	**22** (33%)	**67** (61%)	**4** (13%)	**12** (25%)

3.6 Robustness checks

To test robustness, the pooled main effects model was estimated for different subsamples of survey 1: gender; age group; health and relationship status (those with $X \leq X^*$ vs the rest); depression status; life satisfaction (7 and below vs 8 and above); personal spending (below £10,000 vs the rest); and WTP (positive vs zero). In addition, whether the respondent selected key statements in the feedback section ("confident about my answers"; "not sure about my answers"; "interesting exercise" and "too many tasks"), and the proportion of negative statements were also used. Overall, the results are highly stable. Extreme problems in physical functioning, extreme pain, extreme depression, and having relationships are always strongly significant (typically $p < 0.001$). Some problems in functioning and some pain are never significant. This may suggest that (with the possible exception of personal spending) preference heterogeneity across individuals is not very strong.

Some depression and personal spending fluctuates between significance and non-significance. For example, some depression is not significant for those with zero WTP in Version IU ($p = 0.344$) but it is significant at 5% for all other subgroups. Or, personal spending is clearly not significant for males in Version IU ($p = 0.746$), marginally non-significant for females in Version SW ($p = 0.091$) and males in Version SW ($p = 0.120$) but highly significant for females in Version SW ($p < 0.001$). (Full details available on request.)

3.7 Individual utility vs social welfare

The model that pooled across the two versions and included interactions between the main effects variables and a version dummy using survey 1 resulted in five significant interaction coefficients ($p < 0.05$) out of eight. The results suggest that relative to Version IU where respondents were asked to rank the hypothetical lives on the basis of what they would prefer for themselves, in Version SW where (a different sample of) respondents were asked to rank the hypothetical lives in terms of deservedness for public support, respondents gave relatively more weight to personal spending ($p = 0.08$) and less weight to health problems, which explains the difference in equivalent consumption in Table 7. However, the design of the sequential ranking exercise did not take into account to conduct this analysis. The model results in three interaction terms (some pain x version, some depression x version, and relationships x version) to be omitted because of collinearity. (Full details available on request.)

3.8 How respondents in Survey 2 perceived the survey

3.8.1 Overall clarity and difficulty of the tasks

Regarding the overall clarity of the survey, respondents from the qualitative interview Survey 2 typically stated that the survey as a whole was clear, though a certain level of bias may have been caused here by the presence of the interviewer. The main choice tasks in section (f) were deemed by many to be clear.

In terms of the difficulty of the questions, some respondents commented on the difficulty in reporting their income and spending, in particular at the household level, in section (c). While most respondents knew what they (and their partner, if relevant) earned, some respondents struggled to work out what the household spending was because they did not have the overall financial responsibility of the

household. The sole earners found these easy. Some respondents knew exactly what their income and spending were each month and annually, whereas others were less certain. Those respondents who knew exactly sometimes mentioned that this was due to a lack of savings.

Others stated that the choice task in section (f) was a "challenging" task. Some said that they "weren't great with numbers". Some respondents in Version SW also mentioned that they found it difficult to imagine themselves in the different life scenarios as they lacked any personal experience of the issues, while others used their personal experience when approaching the scenarios.

3.8.2 The number of options offered in multiple choice questions

Section (c) asks respondents to identify one of several amounts indicated that comes closest to their household income, household spending and personal spending: there were five, five and ten options offered, respectively. When asked about their views on this set up, there were three different views. First, some respondents were happy with the number of the options they could choose from and the amounts indicated in each. Second, some respondents struggled to select one of the options and thought that the gaps, in particular between the two highest amounts, were too large. Third, some respondents thought that the income amounts offered corresponded to the pay grades of different kinds of jobs, and therefore found it easier to select the appropriate amount.

Section (e), question 2.2 (WTP2) offers 11 options to indicate the proportion of personal spending (0%, 10%, … 100%) that could (in effect) be traded off against going from X to X^*. Some respondents thought that there were too many options.

3.8.3 Intrusiveness

When asked whether they perceived the survey as intrusive, none of the respondents thought that this was the case. However, some believed that other people may perceive questions relating to income as intrusive (but not the other attributes).

3.9 How respondents in Survey 2 interpreted the various components

3.9.1 Self-reported personal spending

There was some confusion regarding the concept of personal spending. Regarding section (c) which elicits personal spending via household income and household spending, respondents found the concept of spending unclear. Firstly, "personal spending" was not always understood to include all expenditure, including bills, rent and mortgage.

> "*when I first read it, I assumed it was going to be, you know, for stuff on yourself, like going out, clothes, you know personal things*" (R7).

Secondly, respondents sometimes found it difficult to determine their own share of the household spend. This was because they regarded the household expenditure as joint, thus not decomposable to individual shares, especially where they had children (or pets - some asked for clarification on who was included in the household). One approach was to simply divide household spending with the number of people (or adults) in the household.

3.9.2 WTP to go from X to X^*

Regarding the two WTP questions in section (e) which both start with "Suppose your current health and relationship problems went away…", some respondents found difficult as they had previously answered that they had no relationship or physical problems.

WTP1 (question 2.1) continues "… but your spending was cut to half its current level, would you say you would be: better off / about the same / worse off". This was confusing to respondents who were thinking of the impact of improved health on income. For example:

> "if *my health problems went away, then I could get like a full-time job.*" (R15)

Others found the wording convoluted.

> "*Yeah, actually, they could have just cut to the chase and just said, "Does money make you happy? maybe*". (R9)

The question that caused the most serious confusion was WTP2 (question 2.2), which continued: "… what would be your personal spending level that would make you feel as satisfied with your life as during the past 12 months?" Some were puzzled to find that spending should go down even though other problems went away, for example:

> "*If my current health situation was better, then I would say that I would spend more money, not less money*" (R13).

Others assumed that reduced spending implied increased saving (and therefore, a desirable thing).

> "*oh actually, I would be better off because it would be good to kind of save more money than I'm saving at the moment*" (R16).

3.9.3 Conceptualising the attributes of the choice tasks

The five attributes of wellbeing used in the choice tasks in section (f) are discussed below.

Many respondents ranked *physical functioning* as less important for wellbeing than mental health and relationships, but some did regard physical functioning the most important:

> "*I think for me it's, it was always about looking at the problems with physical functioning*" (R9)

Respondents were divided on the issue of *pain*. Some regarded it as having a seriously negative effect on wellbeing:

> "*I'm a big baby when it comes pain*" (R1)

Others did not see it as a major factor:

> "*everybody has pain from time to time, so that's not a major, um, problem*." (R5)

For many respondents, it was important to avoid *depression*, which was seen as very debilitating and difficult to get support for.

> "*having extreme depression would be the most, um, scary for me*" (R17)

Having good *relationships* was for many respondents the most important part of wellbeing. They often mentioned that relationships would help you deal with any of the other problems, implying possible interactions. Without good relationships, all other aspects of wellbeing were seen as considerably more difficult to cope with. For example:

> "*no matter how much money you have or no matter what great health you're in, if you haven't got friends and family, then life's pretty miserable, to be honest*." (R15)

"I think I'd always rank, um, kind of good mental health and personal relationships above everything else I think." (R10)

Of the three *personal spending* levels used in the choice tasks (£6,000 p.a. / £20,000 p.a. / £41,000 p.a.), many respondents regarded £20,000 as "enough" to live comfortably on.   While some regarded £6,000 as extremely low, others thought this was already enough, possibly because they were not considering utility bills and rent/mortgage.

Many respondents stressed that personal spending was the least important item, or it was the last attribute that they looked at when they ranked the imaginary life scenarios.

"I didn't actually really look at the personal spending at all, really." (R8)

"I think if you've got your health it doesn't matter what money you've got really."(R11)

However, other respondents acknowledged the relevance of some money to wellbeing.

"it's easy for people to sort of say that, oh, you don't really need money to be happy, kind of thing, and, and that's true to a degree, but certainly in today's society there's so much, you're limited a great deal by not having enough money."(R15)

"I do know people who are a lot more wealthy than me who I don't think are as happy. Um, but I also think that, you know, we're quite happy but a bit of money would make us even happier." (R17)

3.9.4 Conceptualising the hypothetical lives

The hypothetical lives were intended to last with no change, as given, for an unspecified duration. However, some questioned the independence across the attributes and how an attribute may affect another:

"even relationships, um, you know, you tend to argue more when you've got less money." (R12)

"actually, if you had no good relationships, you know, perhaps you would be depressed." (R17)

Some respondents considered how, over time, one may adapt to problems in some attributes of wellbeing:

"[I] have kind of very moderate chronic pain from a back problem. But you know, I – I sort of – I adjust to it and I – I don't find it that bad".   (R6)

Others introduced exogenous factors such as medical treatment:

"pain can be managed with pills." (R1)

3.10 How respondents in Survey 2 approached the WTP and choice tasks

3.10.1 Trading off between health and relationships vs money

In contrasting health and relationships against personal spending, some respondents thought that health and money were not commensurable:

"I don't think really you can, you can compare the two, health and wealth, really." (R8)

However, many would consider giving up money for better health:

"you would quite willingly give up quite a substantial amount of that [personal spending] to not have some or all of these problems." (R13)

"I'd rather not have much money really [than have depression]." (R15)

Others implicitly accepted the commensurability, but did not think the trade-off worthwhile, because with less money…

"I wouldn't have the, the freedom to, to, you know, to do what I can do now." (R4)

3.10.2 Approaches to the choice tasks

Respondents took a range of approaches (not necessarily mutually exclusive) to complete the choice tasks in section (f).

One approach was lexicographic – in other words, to construct a ranking of the dimensions of wellbeing and then judge the lives according to how they scored on this ranking. In this approach, dimensions of wellbeing were seen to some extent in isolation. For example:

"I've prioritised what I thought was important to me out of those four [attributes] and then looked across [the lives] to see which ones had that one." (R1)

Other approaches were to work backwards by eliminating the lives that the respondent liked the least, or to try to avoid extremes on any of the dimensions as far as possible.

Another approach was to allow the attributes of wellbeing to interact: thus to see across the attributes as a whole and make a judgment on how they would affect one another. These respondents often held that the attributes could not be judged in isolation from each other.

"That's only bad if I don't have this. So yeah, you you think of them as separate things, but then you realise you can't, you do have to look at them all at once." (R2)

"depression isn't so much of a big thing for me, but if physically couldn't do that [go out and do things], then depression would be a problem." (R9)

3.10.3 Trading and ranking of wellbeing items

When comparing across hypothetical lives in the choice tasks in section (f), respondents noted that the items were difficult to compare:

"If you had some more pain, less pain but, um, more extreme depression, it was difficult to weigh it up." (R11)

Some respondents were willing to trade physical functioning or pain for more personal spending, but less so their mental health or their relationships.

"the money did come into it but it was depend', it was more sort of a trade-off with the pain and the physical problems" (R14)

Some respondents seemingly changed their approach as they progressed, which they sometimes reflected upon themselves. They stated that they started by looking at spending levels, but ultimately decided that it was the least important factor for wellbeing.

3.10.4 The effect of perspectives

Many respondents given Version SW held that they approached it from their own personal perspective:

"I did think about it as if it was me and so I used very personal preferences." (R6)

"well, you know, that's the way I would rank it so that's the way I'd kind of see it kind of thing." (R10)

Other respondents reflected upon the difficulty in applying their own preferences, saying that it was a subjective exercise and that their preferences could not necessarily be universalised.

"I think that's just different people value different things." (R3)

Some noted how the perspectives may affect their responses to the choice tasks. For example, the relative importance of spending may increase under Version SW:

"I think I would probably value money more for other people because I do think that – that people should have enough money to live on. And what other people regard as enough to live on isn't necessarily what I regard as enough to live on, for myself, for my own personal spending." (R6)

On the other hand, another respondent had the opposite view:

"I think maybe when it's other people, you're a bit more likely to say, oh, you should want to have your health and you shouldn't worry about the money. But when it's yourself and you actually know how much you need to get by, you perhaps think that a little bit less, which is interesting." (R14)

3.10.5 Timeframe

Respondents said that they either did not have a timeframe in mind at all, or that they thought of the situations as permanent. Some did, however, say that a different timeframe would have affected their answer.

4. DISCUSSION

This paper reports on an informal attempt at operationalising equivalent consumption at the individual and societal levels using two methods: through individual WTP based on the respondents' current situation and through choice experiments using hypothetical scenarios. Two convenience samples were drawn from non-academic non-research staff at the University of Sheffield, to take part in an online survey or a face to face interview. Almost 6,300 e-mail messages were sent, resulting in 96 accesses to the online survey with 52 completions (Survey 1), and 17 interviews (Survey 2).

Because equivalent consumption is defined as a function of the individual's current situation (X,c), the operationalisation of the concept at the individual level builds crucially on having accurate information on the respondent's current situation. The design of the questionnaire made it clear that asking factual questions effectively is not necessarily more straightforward than asking questions on hypothetical scenarios.

In particular, the qualitative Survey 2 highlighted the difficulties associated with the concept of personal spending. There was confusion about what to include in this and how to derive this amount. Some respondents interpreted this to mean expenditure exclusively for personal leisure beyond household expenditure. The term also needs to convey that it is an indicator of benefit (material standard of living). In future studies, an example to illustrate that the sum of personal spending of two people from the same household may be larger than the level of spending of that household because some components of shared spending are public goods (but not in these words) may be useful: viz. it is possible – and probably reasonable – for both people to each claim 80-100% of the rent as their

"personal spend". At the same time, care needs be taken so that respondents do not interpret the example as the "correct" way to calculate their personal spending.

Furthermore, the term "spending" needs to be replaced or better explained to rule out the interpretation that it is independent of income (ie. a reduction in spending does not imply an increase in savings). The phrasing "better off / worse off" may also be interpreted to mean financial standing, as opposed to overall wellbeing. These point to the importance of clearer instructions and suitable examples.

As expected, the achieved sample is mostly healthy, with moderate income and high life satisfaction. In particular, from the two surveys 24 (47%) and 9 (53%) respondents respectively reported no problems in health or relationships, so in our model their current consumption is their equivalent consumption. However, they may have had health or relationship problems that were not captured by the four questions asked – and therefore follow-up questions to pick up those with non-reported problems would be useful. Around a third of the online sample and half of the interview sample indicated a positive direct WTP to achieve X^*, and the higher proportion in the interview is interesting – this may indicate that they have genuine non-reported problems. Amongst the remaining respondents, with reported problems, there was a substantial inconsistency between the warm up question (WTP1) and the actual question (WTP2) in Survey 1. This proportion was significantly smaller (at 10%) in Survey 2.

The choice experiment models based on sequential ranking did not converge for all respondents, in either survey. Where an individual model did converge, most of the coefficients for the *some* levels were not statistically significant, even at 10%. The presence of the interviewer in Survey 2 is associated with significantly fewer actual and potential inconsistent tasks (but this does not necessarily make it more likely that a model will converge). It is also associated with significantly fewer (zero) respondents thinking they do not want to answer any more choice tasks, with 41% reporting they could answer another five or six choice tasks. However, estimating individual level preferences allowing for non-linearity would be a substantial challenge.

Nevertheless, although the individual models did not typically converge, the qualitative data suggest that at least some people do trade across the dimensions of wellbeing, including consumption. Respondents referred to various approaches including trading, trading with interactions, and lexicographic preferences.

On the other hand, the pooled models for all samples converged, with a significant coefficient for *some depression*, alongside all *extreme* problems. Overall, it appears to be possible to operationalise equivalent consumption through choice experiments, linearly, at the aggregate level. The wording of the concept of "personal spending" will be a crucial challenge. A balanced experimental design that allows for interactions both between X and spending and between the X-dimensions will not be possible with S = 6. The next balanced design requires S = 12, which is probably too large (see above), but a smaller unbalanced design or J =5 may be worth considering.

The two approaches to calculate individual equivalent consumption – the direct method based on individual WTP for X^*, and the indirect method based on predictions from the pooled sequential ranking coefficients – resulted in highly correlated values in both surveys (but note that both are a function of current personal spend). The two methods have different pros and cons. The direct WTP method builds on one explicit question on gains in X to X^*, while the sequential ranking involves a number of relatively obscure tasks involving losses in X from X^*. Firstly, there may be gain-loss asymmetry. Secondly, respondents may like the direct WTP method more because it is simpler. However, since the analysis of direct WTP takes individual responses at face value, it is more susceptible to error, while the econometric analysis of choice experiment data assumes observations include measurement error. Some of the respondents in Survey 2 indicated that the direct WTP

method was confusing, because they assumed better health and relationships will lead to higher income and thus higher spending. In other words, while the question was asking for the marginal rate of substitution along the iso-utility curve, these respondents were thinking about the marginal rate of transformation along the production frontier.

In addition, the direct WTP approach does not require a specification of the individual utility function, whereas the indirect method needs a specific individual utility function to model the choice experiment data. Furthermore, in order to model a possibly concave utility function, the direct WTP method does not allow this at the individual level, at least if it only compares the actual situation of the individual with the reference situation.   In that case, the best it can do is to parameterise and estimate preference differences between groups. In principle, it would be possible to extend the direct WTP approach to include more hypothetical situations, e.g. intermediate combinations of X. The richer information obtained in that way could be less reliable, however, precisely because this approach would involve comparing two hypothetical situations.

The indirect method, in theory, can accommodate non-linear specifications at the individual level, but fitting this will be a clear challenge. Pooling across respondents to fit an aggregate model is possible under both approaches, especially with larger sample size. The relative merits of a method that relies the respondent's report of their actual situation (X,c) albeit through only one observation against another method that builds on multiple observations per respondent but based on hypothetical situations need to be carefully assessed in terms of efficiency and error.

As mentioned before, there is another, third, approach to calculate equivalent consumption, through estimating a common "happiness equation" (Frey and Stutzer, 2000; Decancq et al., 2015b). Given that there is only one set of observations per respondent, it is not possible to model a happiness function for each individual. Thus, this approach needs to pool across the individuals to estimate a common happiness equation; and then use these to predict the level of (X^*,c') for each individual that would achieve the same level of life satisfaction as that observed under the actual situation (X,c). However, this approach only works if there is sufficient variation in respondents' current situation and life satisfaction. In addition to the low proportion of respondents with $X < X^*$, the happiness equation for the Survey 1 sample resulted in only one statistically significant coefficient (own depression, $p = 0.15$ ordered logit), and therefore the approach was not pursued further (full details available on request).

The two Surveys resulted in some disagreement, and Survey 2 appears to have resulted in better quality data. For example,

- the correlation between per capita spending and personal spending is higher in Survey 1 than in Survey 2;
- a significantly higher proportion of respondents in Survey 1 were logically inconsistent between WTP1 and WTP2;
- more respondents in Survey 1 gave inconsistent tasks in sequential ranking;
- amongst the individual level regression coefficients, the only statistically significant coefficient with the "wrong" sign was found in Survey 1; and
- more respondents in Survey 1 fed back that they "got tired half way through" and did "not want to answer any more of these tasks".

However, the presence of the interviewer in Survey 2 did not make it more likely for the individual level regression model for the sequential ranking to converge, and the coefficients of the pooled regression models are similar to each other across the two surveys.

Regarding the two survey versions, corresponding to the personal individual utility (IU) and the societal (SW) perspectives, there is mixed evidence. On the one hand, the quantitative data, especially in

Survey 1, resulted in differences that suggest people have different preferences by which to assess their own wellbeing and others' wellbeing in a way that underestimates the monetary value of the health of others. On the other hand, several respondents in Survey 2 stated that they answered the SW version as if it was about themselves. Overall, the values of predicted equivalent consumption varied little between the coefficients based on Version IU and those based on Version SW.

5. CONCLUSION

Equivalent consumption is one possible preference-based measure of individual well-being. Trying to operationalise it fits into a more general objective of formulating a measure of individual well-being that goes beyond health and also takes into account other important life dimensions, such as personal consumption or the quality of social relations. Such a broader concept of well-being would make it possible to come closer to the dominant justice opinions in society, which are not restricted to health only. As Hausman (2007) emphasizes, "a state of affairs in which those who are otherwise worse off are healthier than those who are otherwise more fortunate is *more* just rather than less just than a state of affairs which is exactly the same except that health is equally distributed" (p.50). Moreover, a broader concept of well-being also makes it possible to broaden the scope of health economic evaluations to include in a coherent way budgetary restrictions, i.e. the trade-off between consumption and health, and distributional weights (see, e.g., the application in Samson et al., 2018). As for the latter objective, preference-based measures have the advantage that they do no rely on interpersonal comparisons of subjective utility. On the other hand, the main challenge with such measures is the identification of (preferably individually differentiated) preferences. This is not a trivial task.

This paper reports on a pilot study in two parts that aimed to operationalise equivalent consumption. The study should be regarded as an exploratory exercise to inform future studies, rather than as a source of plausible estimates for equivalent consumption for policy. The quantitative exercise found that, using sequential ranking tasks administered online, we can quantify equivalent consumption linearly through choice experiments at the aggregate level. The coefficients have the expected signs and thus satisfy minimum face validity. Doing so at the individual level, however, is significantly more challenging: individual models for many respondents did not converge, even assuming linearity. Qualitative evidence suggests that the key components of the survey needs further refinement to avoid misunderstanding. There is limited evidence (based on small numbers) that, on the one hand, the quality of observations from online surveys may be inferior to those obtained from face to face interviews, while on the other hand, this may not have a tangible effect on the quantitative parameters estimated.

REFERENCES

Choice Metrics (2012) Ngene [software for experimental design]

Decancq K, Fleurbaey M, Schokkaert E (2015a) Inequality, income and well-being. In: Atkinson A, Bourguignon F, (Eds.) *Handbook of Income Distribution, Vol. 2A*

Decancq K, Fleurbaey M, Schokkaert E (2015b) Happiness, equivalent income and respect for individual preferences. *Economica*. 82, 1082-1106.

Dolan P, Olsen JA, Menzel P, Richardson J (2003) An inquiry into the different perspectives that can be used when eliciting preferences in health, *Health Economics*, 12: 545-551

Fleurbaey M, Luchini S, Muller C, Schokkaert E (2013), Equivalent income and fair evaluation of health care, *Health Economics*, 22:711-729

Fleurbaey M, Schokkaert E (2012), Equity in health and health care, Pauly MV, McGuilre TG, Barros PP eds, *Handbook of Health Economics Volume Two*, North Holland

Frey BS, Stutzer A (2000). Happiness, economy and institutions. *The Economic Journal*. 1;110(466):918-38.

Hausman D (2007) What's wrong with health inequalities? *Journal of Political Philosophy.* 15, 46-66.

Louvier JJ, Street D, Burgess L, Wasi N, Islm T, Marley AAJ, (2008), Modeling the choices of individual decision-makers by combining efficient choice experiment designs with extra preference information, *Journal of Choice Modelling*, 1: 128-163

Office of National Statistics (2014a), Tables 14a, income, tax and benefit data by income quintile for all households, from The Effects of Taxes and Benefits on Household Income, Historical Data, 1977-2012/13 Release, http://www.ons.gov.uk/ons/rel/household-income/the-effects-of-taxes-and-benefits-on-household-income/historical-data--1977-2012-13/summary--historical-data--1977-2012-13.html [accessed 27 August 2014]

Office of National Statistics (2014b), Quarterly National Accounts - Households & Non Profit Institutions Serving Households Sector, http://www.ons.gov.uk/ons/datasets-and-tables/data-selector.html?cdid=NRJS&dataset=qna&table-id=J3 [accessed 27 August 2014]

Ritchie L, Lewis J, eds. (2005) *Qualitative research practice: A guide for social science students and researchers*. London: Sage.

Samson A-L, Schokkaert E, Thébaut C, Dormont B, Fleurbaey M, Luchini S, Van de Voorde C (2018) Fairness in cost-benefit analysis: a methodology for health technology assessment. *Health Economics.* 27, 102-114.

Sen A (1985), *Commodities and Capabilities*

Tsuchiya A, Watson V (2017) Re-Thinking 'The Different Perspectives That can be Used When Eliciting Preferences in Health'. *Health Economics*.

APPENDIX: Example screenshots of the choice tasks

Part 3: your views on imaginary life situations (2)

In this section of the questionnaire we ask you to choose form a set of imaginary life situations, described using the items you have seen in Part 1.

Below is an example of four imaginary life situations. It shows that Life A, in this example, means no problems with physical functioning, no pain, no depression, no relationships with relatives and friends, and an annual personal spending of £6,000.

Lives B, C, and D each have different combinations across the same items. We will ask you to think about imaginary lives like these and ask you to tell us which of these is the best, the second best … and so on.

	Life A	Life B	Life C	Life D
Problems with physical functioning	None	Extreme	Some	Extreme
Pain	None	Extreme	Some	No
Depression	None	None	Moderate	Extreme
Have good relationships	No	Yes	Yes	No
Personal spending	£6,000	20,000	£6,000	£41,000

Task 1

Please take a look at the table below that summarises four imaginary lives (A to D) that you could live in. Please think about which of these lives you think are the best and the worst. There are no right or wrong answers.

	Life A	Life B	Life C	Life D
Problems with physical functioning	None	Some	Extreme	None
Pain	None	Some	Some	Extreme
Depression	Moderate	None	Extreme	None
Have good relationships	No	Yes	No	Yes
Personal spending	£20,000	£20,000	£6,000	£6,000

*

	Life A	Life B	Life C	Life D
Which life do you prefer the most	○	○	○	○
Which life do you prefer the second	○	○	○	○
Which life do you prefer the third	○	○	○	○
Which life do you prefer the least	○	○	○	○