Article:
Ta, A., Van Landeghem, B. and Tsuchiya, A. (2024) Eliciting public preferences across health and wellbeing dimensions: an equivalent income value set for SIPHER-7. Health Economics, 33 (12). pp. 2723-2741. ISSN 1057-9230

https://doi.org/10.1002/hec.4890

Received: 27 April 2023 | Revised: 24 July 2024 | Accepted: 2 August 2024

DOI: 10.1002/hec.4890

RESEARCH ARTICLE

Health Economics WILEY

Eliciting public preferences across health and wellbeing dimensions: An equivalent income value set for SIPHER-7

An Ta[1] | **Bert Van Landeghem**[2] | **Aki Tsuchiya**[3]

[1]School of Medicine and Population Health, The University of Sheffield, Sheffield, UK

[2]School of Economics, The University of Sheffield, Sheffield, UK

[3]School of Medicine and Population Health and School of Economics, The University of Sheffield, Sheffield, UK

Correspondence
An Ta, School of Medicine and Population Health, The University of Sheffield, 30 Regent St, Sheffield City Centre, Sheffield S1 4DA, UK.
Email: a.t.ta@sheffield.ac.uk

Funding information
The University of Sheffield School of Economics; Systems Science in Public Health and Health Economics Research Consortium, Grant/Award Number: MR/S037578/2

Abstract

The call for "health and wellbeing in all policies" requires a preference-based measure that collapses multi-dimensional health and wellbeing into a single index, such as equivalent income. We aim to elicit preferences of the UK general public to estimate a value set for a suite of seven commonly used wellbeing indicators including health, income, and other dimensions, in terms of equivalent income. Secondly, we examine heterogeneous preferences by gender, by age, and by income. Thirdly, we explore the stability of preferences, since the survey took place amid the pandemic, possibly affecting preferences over health and wellbeing. Effects of attrition and of time are distinguished. Data were collected online across two waves using Discrete Choice Experiments through an internet panel ($N1 = 3362$; and $N2 = 3357$). The regression coefficients for all the ordered attribute levels have the expected sign, are significant, and ordered. Equivalent income was found to vary up to 10% by gender and by age (both significant) and 4% by income (not significant), while the effect of time was up to 16% (significant). The study facilitates the calculation of overall wellbeing in terms of equivalent income based on the preferences of the UK public, where the relevant wellbeing indicators are available.

KEYWORDS
discrete choice experiment, equivalent income, stated preferences, wellbeing

JEL CLASSIFICATION
I1 Health, I3 Welfare, Well-being, Poverty

1 | INTRODUCTION

There is a call for a move from "health policy" to "health and wellbeing in all policies" (e.g., WHO, 2014), arising from the growing recognition that wellbeing is multi-dimensional, requiring an inter-sectoral approach to public policy. Currently, health economists conduct evaluation of health interventions using health outcome measures (e.g.,

EQ-5D). However, the evaluation of policies such as changes to welfare payment entitlements or minimum wage would need to consider their effects across multiple dimensions of wellbeing besides health, including income, employment, housing, feeling safe, etc (Capéau et al., 2020; Stiglitz et al., 2009), and this leads to a need for a preference-based outcome measure that collapses multiple dimensions of wellbeing into a single index. "Equivalent income" (Fleurbaey, 2005, 2006, 2009; Fleurbaey & Gaulier, 2009; Fleurbaey & Schokkaert, 2011) is a preference-based monetary measure of wellbeing that adjusts income for non-income dimensions of life, considering people's preferences on how important those different dimensions are to them. In other words, instead of taking income as the sole measure of wellbeing, it penalises income for non-income dimensions of wellbeing at sub-optimal levels, based on people's preferences (see Section 2.1 below for more details). This is similar to the concept of the Quality Adjusted Life Year that, instead of taking survival as the sole measure of health, penalises survival for sub-optimal health-related quality of life, based on average preferences of the respondents. While the concept of equivalent income was originally developed to capture individual preferences, the equivalent income value set we estimate deviates from this, as we will aggregate across individual preferences, to report average equivalent incomes for different multi-dimensional outcome states.

The first aim of this study is to estimate an equivalent income value set through an online Discrete Choice Experiment (DCE) survey using a suite of seven wellbeing indicators called SIPHER-7 (Tsuchiya & Wu on behalf of the SIPHER Consortium, 2021). The seven wellbeing indicators have been selected so that they can be linked to survey questions used in the UK Household Longitudinal Study—UKHLS, also known as Understanding Society (University of Essex and Institute for Social and Economic Research, 2020) (See Appendix 1). The parameters estimated from the DCE model allow us to calculate equivalent income for all SIPHER-7 profiles, and thus any UKHLS respondent.

The existing empirical literature on modeling equivalent income has mostly relied on the revealed preference approach and estimates from subjective wellbeing (or happiness) regressions. These use national surveys such as the UKHLS, the German Socio -Economic Panel (GSOEP), the Flemish survey (SONAR) and Russia Longitudinal Monitoring Survey (RLMS-HSE) to calculate how much income an individual could trade in for better non-income dimensions while maintaining the same happiness level (see Decancq et al., 2015; Fleurbaey, 2009; Fleurbaey & Blanchet, 2013).

On the other hand, relatively little work has been carried out using the stated preference approach, and we are aware of three, of which two used contingent valuation. Fleurbaey et al. (2013) surveyed 542 members of the public in Marseilles, France, and used contingent valuation across income and health to model preferences at subgroup level. Respondents were asked to think about a hypothetical scenario in which they could trade-off income for perfect health and required to answer the income level (equivalent income) that they would be equally happy to have if they had no health problems. Capéau et al. (2020), which reports on the MEqIn project, is the largest of the three studies. The project conducted a large-scale survey of the Belgian general public, interviewing more than 3000 adults from over 2000 families. The survey examined several dimensions of individual wellbeing including income, consumption and affordability, health, work status, housing, and time-use. Furthermore, it asked respondents for their willingness to pay (WTP) to achieve the best levels in the non-income dimensions. We are aware of only one study that has used a DCE to model equivalent income. Abasolo et al. (2018) is a small-scale exploratory study of a convenience sample ($N = 52$ for the quantitative analysis) in Sheffield, UK, and used a DCE based on sequential ranking to model individual-level preferences across Physical functioning, Pain, Depression, Close relationships, and Personal spending.

Our second aim is to examine heterogeneity in preferences across observable population groups by modeling the DCE data separately for different subgroups defined by gender, age or equivalised household income. We focus on heterogeneous preferences by observable characteristics, since, if information on preference heterogeneity is to be of use to policy makers, arguably, they need to be based on observable characteristics. However, to control for unobservable individual level heterogeneity mixed logit estimations are used.

The third aim is to explore the stability of preferences over time. We repeated the same DCE survey for a second time, roughly a year later, to control for any seasonal effects. The first DCE survey took place in the autumn of 2020, in the midst of a pandemic and associated social restrictions. This might have made people more concerned about the health-related dimensions of wellbeing and/or about feeling lonely. Thus, if the SIPHER-7 value set is to be used to evaluate policy outcomes beyond the pandemic, it is important to establish that the preferences elicited are not

transient. One approach to examining the stability of preferences over time is to survey the same respondents again at some later time. However, such panel samples may overestimate the stability of preferences due to a selection effect, if those with unstable preferences and/or those living unstable lives are more likely to drop out from later surveys. Furthermore, "panel conditioning" has been observed, where the experience of repeating the same survey itself has an effect on the respondent's responses in future surveys. For example, Van Landeghem (2014, 2019) found that, ceteris paribus, reported satisfaction scores decreased the more often one had been exposed to such questions and Fisher (2019) found that more experienced respondents provided higher-quality income data. Another approach is the repeated cross-section, to recruit fresh but interchangeable samples from the same population using the same sampling method so that observed differences in preferences can be attributed to changes in preferences within (a subgroup of) society over time and not to differences in the sample stemming from attrition or having more survey experience. However, this approach needs to assume that unobservable features of the population have remained stable in the meantime. Our second wave was a hybrid of the panel approach and the repeated cross-section approach. To summarise, this paper reports on a stated preference study to calculate an equivalent income value set for SIPHER-7, by conducting two waves of an online survey each with over 3000 respondents in the UK using a DCE. The value set will allow analysts to convert the effects of different social and public policies into a preference-based single-index monetary metric of wellbeing, and to inform inter-sectoral policy decision-making. We examine preference heterogeneity across observable subgroups. By conducting two waves of the survey with both repeating and fresh respondents in Wave 2, we examine the stability of preferences across the survey waves, distinguishing between the effects of attrition and of time.

2 | METHODS

2.1 | Equivalent income

Assume that individual (U) wellbeing is a function of a finite number of observable indicators, including income. Equivalent income (Y') is defined as the hypothetical amount of individual income that, if combined with the best levels of non-income indicators (represented by vector $\boldsymbol{X}^*$), is as good as the current situation (Fleurbaey, 2005; Fleurbaey & Schokkaert, 2011):

$$U(X, Y) \equiv U(X^*, Y') \tag{1}$$

If an individual's non-income profile (X) is made up of the best levels across all the non-income indicators, then equivalent income will be the same as this person's current income. On the other hand, if any of the non-income indicators are not at the best levels, the individual's equivalent income will be no higher than the current income level.

Assuming no loss aversion, the difference between current income and equivalent income corresponds to the concept of WTP for an individual at X, Y' to achieve the best non-income profile.

$$\mathrm{WTP}_{(X,Y)} = Y - Y' \tag{2}$$

There are two things to note. First, given the way our DCE was designed with a neoclassical utility function in mind, there is no distinction between the amount of money that an individual is willing to *pay* to go from $\boldsymbol{X}$ to $\boldsymbol{X}^*$ and the amount of money that the same individual is willing to *accept* to go from $\boldsymbol{X}^*$ to $\boldsymbol{X}$, but was not accommodated in our choice design. Thus, what we refer to as the WTP to achieve the best non-income profile is not expected to predict actual measurements of WTP to go from $\boldsymbol{X}$ to $\boldsymbol{X}^*$ since the empirical amount is likely to be subject to asymmetry between gains and losses. Second, DCEs have been widely used to elicit people's WTP for given goods and services outside of the equivalent income literature, where the monetary attribute typically concerns payments (e.g., price of a good, or increased taxation for a public service). This approach could be adapted to measure equivalent income, by including in the DCE a payment attribute to achieve the best non-monetary profile of wellbeing, in the form of general taxation, for example, Subtracting this from the respondent's own income would allow an indirect calculation of equivalent income.

An alternative, developed in Abasolo et al. (2018) and pursued in this paper, is to use income (as opposed to payment) as the monetary attribute. Here, respondents will be choosing between different life scenarios where income is one of the features that describe them. The approach can expect to avoid focusing on the payment vehicle and allows a more direct elicitation of equivalent income.

2.2 | The discrete choice experiment attributes and levels

The wellbeing indicators of SIPHER-7 consist of: the Effect of Physical Health (hereafter 'Physical health' for short), the Effect of Mental Health ('Mental health'), Loneliness, Household disposable income, Employment, Housing quality, and Neighborhood safety. These indicators and the levels of each were selected to be relevant to a wide range of public policies, through an iterative consultation process within the wider research project involving researchers (with backgrounds in data science, decision modeling, economics, geography, psychology, public health, social policy, etc) and policy partners (decision makers and analysts from three local government bodies), and are reported elsewhere (Tsuchiya, A. & Wu, C. on behalf of the SIPHER Consortium, 2021).

While most of the attributes are self-explanatory, Employment and Household disposable income need some further explanation. The Employment attribute has six non-ordered categories. Respondents are informed that Full-time employed (or 'FT employed' hereafter) and Part-time employed ('PT employed') include being self-employed and being on maternity, parental, sick, or furlough leave. Job-seeking is qualified as "looking for employment." Hereafter in this paper, Taking care of a family member with chronic illness or disability will be referred to as "informal caregiving," Full-time education/training/apprenticeship as "FT education," and Not working and not looking for paid employment as "Not working." For the Housing Quality attribute, attribute levels "Partly true" and "Not true" are referred in the analysis as "Fair" and "Poor" housing respectively.

The Household disposable income had five levels, set as 40%, 60%, 80%, 100% and 120% of a reference household disposable income for the choice design, but presented to respondents in actual monetary amounts. The use of household income as the monetary attribute in a DCE, as opposed to an attribute framed as payments, brings a particular challenge. The range of household income across the population is much wider than the range of realistic amounts that people may be willing to pay for a given good or service. Furthermore, household income of any level needs to be adjusted for household size to be meaningful. To ensure that the respondents can relate to the income levels they see in the DCE, we created 10 DCE variants each corresponding to one of the decile groups of household disposable income in Wave 9 of UKHLS. Reference income was defined as the average household disposable income of the corresponding decile group. Prior to the DCE, respondents were asked to indicate their household disposable income across these decile ranges, and allocated to the corresponding DCE variant, so that the income levels each respondent saw in the DCE were aligned to their own self-reported household disposable income, which was equivalised or household size later in the modeling.

2.3 | The discrete choice experiment design

The DCE model with household disposable income as a continuous variable and interacted with employment involves 25 parameters to be estimated, whereas treating household disposable income as a categorical variable would mean 43. To accommodate the latter comfortably, the DCE consisted of 120 choice tasks, which were divided into 12 blocks of 10 choices.

The design software Ngene was used to generate a D-efficient "partial profile" design assuming scale and preference homogeneity (ChoiceMetrics Pty Ltd, 2018) that allowed at least two out of the six non-income dimensions in each choice to be tied across the alternatives. We started from an efficient design with 120 pairwise choice tasks and the identification of all parings of these 240 scenarios that had two out of the six non-income attributes tied (there are 3600 such pairings). From this set of candidate choice tasks, 120 parings at a time were drawn 10 times to generate 10 partial profile designs, avoiding choice tasks with identical alternatives and dominant alternatives. The final design was one of these 10 designs with the lowest D-error. D-efficiency of the selected partial profile design fell by less than 10% relative to the original full profile design.

2.4 | The qualitative and quantitative pilots

A qualitative pilot aimed to examine how respondents interpret the wording used in the survey and to obtain feedback on the length and difficulty of the survey. A convenience sample ($N = 10$) of university estates staff with no graduate degree was asked to complete the online survey within a video interview with one of the authors (ATT). After this qualitative pilot, changes were made to the survey (details available in the Appendix 7). The revised design was then fielded through a quantitative pilot with 100 respondents in August 2020 using the same internet panel and sampling frame as the main survey. To avoid having to set up 10 income variants of the DCE survey for the quantitative pilot, those who either reported household disposable income between the first and the fifth deciles or did not answer the self-reported income question were allocated to the third income decile variant and those who reported household disposable income between the sixth and the 10th deciles were allocated to the eighth income decile variant. The quantitative pilot provided the priors of the parameters used in the final DCE design.

2.5 | The survey protocol

For the purpose of this study, the two waves of the survey were set up identically to each other. The online survey started with an *Introduction and Informed consent* questions. These were followed by background questions covering the SIPHER-7 indicators, which aimed to help respondents to familiarise themselves with the DCE attributes. There was also a question regarding household size, gender and age. Respondents were then filtered to one of 10 survey variants corresponding to their household income levels. Those who did not answer the income question were allocated to the variant for the fourth decile.

At the start of the DCE section, respondents were asked to complete three practice questions of increasing difficulty, presented in a tabular format as were the main DCE tasks (see Appendix 2 for the first practice question). The first practice question included five tied attributes out of seven and a dominant option. If a respondent chose the dominated option, they were given the reason why their answer was not correct, and the option to re-do the question or to seek further explanation on the choice task and the tabulated format. Practice 2 also included a dominant option but with four tied attributes. The last practice question had three tied attributes, with no dominant option, while the actual choice tasks had at least two ties.

After completing the practice questions, respondents were directed to the 10 choice tasks, with the reminder that there are no right or wrong answers for these. Respondents were given one of the 12 DCE blocks within each income range variant. Within each block, the order of the 10 choice tasks was randomised for each respondent. The left-right alignment of the choice tasks was fixed, and there were no further embedded tasks to check for logical consistency. Figure 1 below presents an example of a choice task.

Once respondents completed the choice tasks, they were presented with a feedback section where they were asked to select from positive and negative statements describing their experience completing the DCE. Following that, respondents were asked questions about their demographics.

2.6 | The two survey waves

The online surveys were hosted on Qualtrics. The respondents were recruited from the PureProfile panel, with age and sex quotas. Respondents from either survey who spent less than 30% of the average completion time of the Wave 1 survey (i.e., less than 5 minutes) were excluded from the analysis. Wave 1 took place in October/November 2020, and Wave 2 was in mid/late November 2021.

Our second wave was a hybrid of the panel approach and the repeated cross-section approach. All respondents to the first wave were re-invited to form the panel sample and to participate in the second wave. Subsequently, the relevant age/sex quotas for the Wave 2 survey were made up by recruiting a further "refreshment" sample recruited from the same pool to form the repeated cross-section sample.

Thus, the DCE data can be broken down into four mutually exclusive sets of observations:

- Panel Wave 1 observations, referred to as P1
- Panel Wave 2 observations, referred to as P2

Please imagine ***a year of your life***, to take place in the next three to five years. If a year of your life in the near future could be like A or B, which would you prefer?

Everything that is not mentioned is the same for A and B. ***After one year, you will return to your normal life***.

Domain	Life scenario A	Life scenario B
You accomplish less because of physical health:	Some of the time	Most of the time
You accomplish less because of emotional problems:	All of the time	Some of the time
You feel lonely and isolated from others:	None of the time	Some of the time
Disposable income of your household is:	£1,040/month or £260/week	£1,390/month or £350/week
Your employment situation is:	Job-seeking	Not working, and not looking for paid employment
Your home is in a reasonable state of repair, has reasonable facilities (cooking/washing) and provides reasonable warmth:	Partly true	Partly true
You are concerned about the safety of the neighbourhood you live in:	Hardly ever	Hardly ever

I prefer...

◯ Life scenario A ◯ Life scenario B

FIGURE 1 An example of a choice task.

- Cross-section sample Wave 1 data only, referred to as C1 (i.e., the attrition sample only observed in Wave 1)
- Cross-section sample Wave 2 data only, referred to as C2 (i.e., the refreshment sample only observed in Wave 2)

The first aim, to estimate an equivalent income value set for SIPHER-7, was addressed using a "value set sample" that included all respondents once, namely either P1 or P2, combined with both C1 and C2. The choice between P1 and P2 is arbitrary, and there is little difference either way (See Appendix 5 for regression results). Subsequently, this value set sample was broken up into subgroups by gender, age or household income to answer the second aim, to examine preference heterogeneity across these observable characteristics. The analyses serving the last aim, to explore differences between those who will attrit and those who will stay, and the stability of preferences over time amongst first-time respondents, used combinations of the four sets of observations, which are explained below (2.7.4) in more detail.

2.7 | Data analysis

2.7.1 | The discrete choice experiment data

The model to estimate is:

$$U_i = \beta_1 Ph + \beta_2 Mh + \beta_3 L + \beta_4 E + \beta_5 H + \beta_6 S + \gamma_1 lnY^i + \varepsilon_i \quad (3)$$

where the notations are the same as those in Table 1, except for lnY^i, the logarithm of equivalised disposable income, for each individual respondent. The levels of the non-income attributes except for Employment were treated as categorical and ordered, with the best level as the baseline, so that the coefficients were expected to be negative with an increasing effect. While the categories of the Employment attribute are not logically ordered, the analysis

TABLE 1 The attributes and levels used in the DCE.

Attributes	Attribute levels
Physical health—Ph	Five levels of frequency:
- In the past 4 weeks, have you accomplished less than you would like with your work or other regular daily activities as a result of your physical health...	1. **None of the time**/2. Little of the time/3. Some of the time/4. Most of the time/5. All of the time
Mental health—Mh	Five levels of frequency:
- In the past 4 weeks, have you accomplished less than you would like with your work or other regular daily activities as a result of your emotional problems (such as feeling depressed or anxious)...	1. **None of the time**/2. Little of the time/3. Some of the time/4. Most of the time/5. All of the time
Loneliness—L	Three levels of frequency:
- You feel lonely and left out from others...	1. **Hardly ever**/2. Some of the time/3. Often
Household disposable income—I[a]	Five levels (40%, 60%, 80%, 100%, 120%) of a reference disposable income level.
- Monthly (or weekly) household income (i.e. income from wage, benefits, pensions, etc, after deducting your tax, national insurance, any occupational pension contributions, and after deducting your rent, mortgage payments or other housing costs).	
Employment—E	Six non-ordered categories:
- This is about your main daily activity. If you are currently on maternity/parental/sick or furlough leave, you count as being employed. (Please choose the option that best describes your main daily activity).	1. **Full-time employed** 2. Part-time employed 3. Job-seeking 4. Full-time education/training/apprenticeship 5. Taking care of a family member with chronic illness or disability 6. Not working, and not looking for paid employment
Housing quality—H	Three levels of quality
- Your home is in a reasonable state of repair, has reasonable facilities (cooking/washing) and provides reasonable warmth...	1. **All true**/2. Partly true/3. Not true
Neighborhood safety—S	Three levels of frequency of feeling concerned
- You are concerned about the safety of the neighborhood you live in...	1. **Hardly ever**/2. Some of the time/3. All of the time

Note: Reference levels/the best levels of categorical attributes for the DCE model are in bold.

Abbreviation: DCE, discrete choice experiment.

[a]Household disposable income (I) is treated as continuous. Both per month and per week values are displayed in the DCE. Only per month values are reported here.

assumed that respondents shared a common ranking, and a pragmatic choice was made to use FT employment as the reference category. The γ_1 coefficient for income is expected to be positive. The specification omits the intercept, since our interest is in the relative size of the coefficients, in the form of equivalent income (see the next subsection).

The DCE model pooled across the 10 income variants and the two waves for the analysis. Household disposable income in the DCE (I) was entered as the logarithm of equivalised disposable income, for each respondent *i*, using their household size based on the OECD square root scale method (OECD, 2008; OECD, 2011). While the DCE choice design assumed a model including interactions between income and employment, given the lower Bayesian Information Criterion of the main effects model, and the additional complexity involved in calculating equivalent income based on the interaction model, a decision was made to use the main effects model as the main model in this analysis. (Results of the interaction model are in Appendix 5).

To control for the repeat observations from each respondent (within a wave) and for individual level preference heterogeneity, we use the mixed logit estimation (Hole, 2007). All non-income attributes were included as random variables with a normal distribution.

2.7.2 | The calculation of equivalent income

Once a DCE model is estimated, the preference parameters can be used to calculate equivalent income for any given SIPHER-7 profile. Let us use vector X^k to represent the six non-income attributes in the model ($k = ph, Mh, L, E, S$ and H). Given the definition of equivalent income (Equation 1 above), based on the main effect model in Equation (3), for any combination of the seven attributes, we have:

$$\sum_{k=1}^{K} \hat{\beta} X^k + \hat{\gamma}_1 \, ln \, Y^i = \sum_{k=1}^{K} \hat{\beta} X^{*k} + \hat{\gamma}_1 ln Y^{i'} \tag{4}$$

where X^{*k} captures the best levels of these six attributes. Given dummy coding, the reference values X^{*k} equal to 0, and therefore equivalent income for any SIPHER-7 profile $\left(X^k, Y\right)$ can be calculated by solving Equation (4) for:

$$Y^{i\prime} = Y^i . \exp\left(\sum_{k=1}^{k} \left(\frac{\hat{\beta}}{\hat{\gamma}_1} \cdot X^k\right)\right) \tag{5}$$

2.7.3 | Subgroup preferences

Subgroup analyses were carried out to explore heterogeneity in preferences. It has been observed that preferences can vary by gender (Decancq & Schokkaert, 2016; Dolan & Roberts, 2002), age groups (Dolan & Roberts, 2002; Tangian, 2005), or income (Quattrociocchi, 2018; Ura & Ellis, 2008). Furthermore, public policies often target subpopulations based on gender, age and income, and therefore these seem reasonable subgroups to explore. We did not have prior expectations regarding the directions of any differences in preferences.

Parameters of a DCE model are on a latent scale and represent scale effects alongside relative preferences between attributes and levels, and thus are not directly comparable across models. Therefore, subgroup preferences were compared in terms of WTP to achieve the best levels in the different attributes and in terms of equivalent income for select SIPHER-7 profiles. Since the money metric is no longer on a latent scale, but is on an absolute scale, this approach by-passes the issue of scale heterogeneity, and the output can be directly compared across subgroups.

2.7.4 | Stability of preferences across survey waves

A preparatory analysis examined the effect of attrition, namely, amongst those who participated in Wave 1, whether the Stayers (who will stay in Wave 2) have different preferences from the Attritors (who will drop out)—this used observations P1 and C1. This analysis is contained within observations from Wave 1, but information from Wave 2 is necessary to identify the Stayers and the Attritors.

The main analysis across waves examined the stability of preferences over time in two ways. (1) The panel approach compared the Wave 1 responses and Wave 2 responses of those respondents who participated twice (i.e., the Stayers)—this analysis uses observations P1 and P2. (2) The repeated cross-section approach compared those in Wave 1 and those who were surveyed for the first time in Wave 2—this uses observations from Wave 1 (P1 and C1) alongside first-time, refreshment, observations from Wave 2 (C2). To control for the different composition of the sample, this analysis is repeated using sampling weights. By employing both approaches, we provide a robust examination of the stability of preferences over time.

For these analyses, the DCE model was fitted separately by subsample and the estimated coefficients were used to calculate the WTP for each attribute level with a starting equivalised disposable income of £1000/month. In addition, equivalent income values were computed for SIPHER-7 profiles with the largest difference in WTP across subgroups.

3 | DATA

3.1 | The respondents

Of the 3623 respondents of Wave 1, 93% took more than five minutes to complete the survey. Of these, 11 individuals were excluded from the modeling as they did not answer the background question on household size, resulting in 3362 respondents included in the analyses. For the Wave 2 survey, 3680 responded, with 97% completing in more than five minutes. Of these, 24 were excluded because of missing household size, resulting in 3533 respondents included in the analysis. P1 and P2 have the same 1185 respondents, while C1 has 2177 and C2 has 2348 unique respondents.

The value set sample based on P2, C1 and C2 consists of 5710 respondents and is largely representative of the UK public in terms of gender and age. Many respondents self-reported the best levels of the SIPHER-7 wellbeing indicators, with 38% of the participants reporting the best levels in Physical health, while 43% of the total with Mental health, 47% for Loneliness, 83% for Housing condition, and 63% for Safety of their neighborhood. The majority were FT employed (39%) or retired (21%). In contrast, regarding Household disposable income, 26% and 17% belonged to the bottom two deciles, respectively. Six percent of the value set sample reported FT employment combined with the best level in the remaining five non-income attributes. Fifty-seven percent of the respondents were either married or living as a couple; 30% had the highest education qualification at degree level or equivalent; and 71% did not have dependent children living with them (see Appendix 3 for details).

3.2 | Feedback from the survey

Overall, most respondents of the value set sample selected positive statements to describe their experience of the DCE. Thirty-six percent of the value set sample indicated that they were confident about their answers, whereas 14% were not sure about their answers. Twelve percent found that there were too many tasks within the survey (see Appendix 4 for details).

4 | RESULTS

The main effects regression model using the value set sample (P2, C1 and C2) results in all the coefficients for the non-income ordered attributes ordered so that worse levels have larger negative impacts. Household disposable income is positive as expected and significant at 0.1%. PT employment is associated with a significant positive coefficient at 1% while other Employment attribute levels have negative and significant coefficients at 0.1%. (The regression coefficients, alongside those from the interaction model, are reported in Appendix 5).

4.1 | Equivalent income and willingness to pay for hypothetical SIPHER-7 profiles

Equivalent income was computed for hypothetical SIPHER-7 profiles, selected for illustrative purposes (see Table 2). Also shown is WTP to achieve the best non-income profile ($\boldsymbol{X}^*$) in absolute terms, and in relative terms as a proportion of equivalised income. The first row, for example, displays a SIPHER-7 profile, with the best levels in all non-income dimensions except for Physical health at its worst level, along with £1000 in equivalised disposable income. The modeled preferences of the value set sample regard this profile to be equally good as an equivalent income of £491/month (viz. the best non-income profile combined with an equivalised income of £491/month), resulting in absolute WTP of roughly £509 per month, or just more than 50% of their equivalised disposable income. The next row illustrates that, given Equation (5) above, the size of equivalent income relative to equivalised disposable income remains constant when disposable income is doubled to £2000 but everything else is kept the same.

The third and fourth rows contrast the effects of having the worst levels in the relatively more personal attributes (health and loneliness) versus the less personal attributes (work, housing and neighborhood). They result in equivalent income amounts of £160/month and £191/month, respectively. And finally, the fifth profile has the worst level in all non-income attributes, combined with an equivalised disposable income of £1000. This profile brings about the lowest

TABLE 2 Absolute equivalent income and relative equivalent income for select SIPHER-7 profiles.

	Physical health	Mental health	Loneliness	Equivalised disposable income (£/month)	Employment	Housing quality	Neighborhood safety	Equivalent income (£/m) [95% CI]	WTP (£/m)	% EI in Y
1	All of the time	[a]	[a]	1000	[a]	[a]	[a]	491.43 [472.45 510.42]	508.57	49.14%
2	All of the time	[a]	[a]	2000	[a]	[a]	[a]	982.86 [944.89 1020.84]	1017.	49.14%
3	All of the time	All of the time	Often	1000	[a]	[a]	[a]	160.36 [150.53 170.18]	839.64	16.04%
4	[a]	[a]	[a]	1000	Informal caregiving	Poor	All of the time	191.07 [176.20 205.94]	808.97	19.11%
5	All of the time	All of the time	Often	1000	Informal caregiving	Poor	All of the time	30.64 [27.28 34.00]	969.36	3.06%

Abbreviations: £/m, pounds per month; CI, confidence interval; EI, equivalent income; WTP, willingness to pay; Y, equivalised disposable income.

[a]indicates the reference level of the attribute.

equivalent income of almost £31/month corresponding to 3% of equivalised disposable income. Note that equivalent income is a way of representing multi-dimensional wellbeing in a single cardinal number, and this number does not represent actual amounts of money that people can (or do) live on (there is no theoretical lower bound).

4.2 | Subgroup analysis using the value set sample

Heterogeneous preferences are examined across three observable characteristics: binary gender, age tertiles, and household disposable income tertiles. The main effects model is estimated for each subgroup. WTP for each attribute level with 95% confidence intervals (CIs) and WTP plots are presented below using the hypothetical equivalised disposable income of £1000/month. If there is no overlap between the CIs across subgroups, this is interpreted as evidence in support of preference heterogeneity between the subgroups within the value set sample at the attribute level. To examine group-level heterogeneity at the profile level, the regression coefficients are used to compute equivalent income for the hypothetical SIPHER-7 profiles with the largest gap between subgroups.

4.2.1 | Subgroups by gender

Of the value set sample, 46% are male and 54% are female. Figure 2 below illustrates the preferences in terms of WTP for each attribute level calculated for the male and the female subgroups separately. The horizontal axis represents the WTP in the negative range (WTP of £100 is represented as -£100). As can be seen, women have a significantly larger WTP than men with regards to the worse levels of Physical health and of Mental health. Women also have a significantly higher WTP for the worst attribute level in Neighborhood safety.

Table 3 illustrates the extent of the differences in preferences across subgroups by identifying the SIPHER-7 profile resulting in the largest difference in equivalent income, where the first row is for the male and female subgroups. As was seen in Figure 2, since females tend to have larger WTP, for each SIPHER-7 dimension, we have picked out the level where the female WTP is the largest relative to the male WTP to calculate the equivalent incomes with CIs. For the SIPHER-7 profile shown, with a baseline equivalised disposable income of £1000/month, male respondents have a significantly higher estimated equivalent income of £186 that is almost double the figure for females (£106). The difference, amounting to 9% of the hypothetical equivalised disposable income, is the largest between male and female equivalent income, which have non-overlapping CIs.

4.2.2 | Subgroups by age

Of the value set sample: 35% are aged between 18 and 40; 33% are between 41 and 57 years old; and 32% are above 58 years old. As Figure 3 shows, across these tertile age groups, the only differences are for some of the Employment categories and Neighborhood safety. Between the youngest age group and the second age group, there is a clear discrepancy in their preferences over "not working" (the younger group has significantly larger WTP) and for the worst level of Neighborhood safety (where the middle-aged group has larger WTP). Preference heterogeneity is observed in the Employment dimension (i.e., 'job-seeking', 'informal caregiving' and 'not working') when comparing the middle and the oldest age groups. By contrast, larger differences in preferences are observed between the youngest and the oldest age groups. The CIs do not overlap for the second to last attribute level ('most of the time') of Mental health, "some of the time" and "all of the time" for Neighborhood safety, "job-seeking," "informal care giving" and "not working" (where the youngest group has larger WTP). In terms of the largest difference in equivalent incomes, Table 3 shows that the SIPHER-7 profile with the largest gap in equivalent income across two age groups depends on the age group pairing. Across the three age group pairings (and all subgroups), the largest gap is observed between the youngest age group versus the middle-age group (10%) with non-overlapping CIs. When comparing the discrepancies between 18-40-year-olds and 41-58-year-olds, the oldest group has an estimated equivalent income of roughly two thirds of the younger group, with non-overlapping CIs. A smaller gap in equivalent income was observed when comparing the middle-age group and the oldest group, with overlapping CIs.

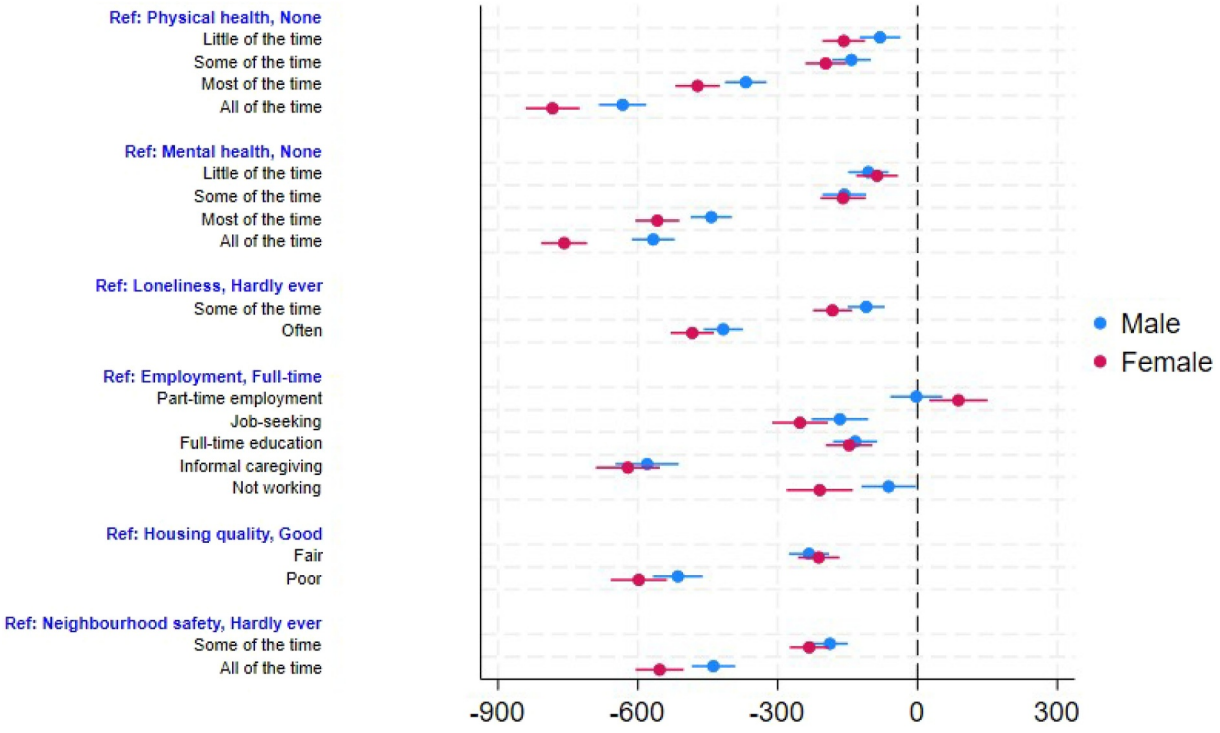

FIGURE 2 WTP for attribute levels by gender groups. The plot was based on a mixed logit model allowing random coefficients (non-income). WTP, willingness to pay.

TABLE 3 Absolute equivalent income and relative equivalent income for select SIPHER-7 profiles between subgroups.

	Subgroup pair	Physical health	Mental health	Loneliness	Employment	Housing quality	Neighbor-hood safety	Equivalent income (£/m) [95% CI]	WTP (£/m)	% EI in Y
1	Male	All of the time	All of the time	Often	Not working	Poor	All of the time	185.66 [167.01 204.32]	814.34	18.57%
	Female							105.93 [91.21 120.64]	894.07	10.59%
2	Aged 18–40	Some of the time	All of the time	Often	FT education	Fair	All of the time	318.35 [272.15 364.55]	681.65	31.84%
	Aged 41–58							221.81 [197.82 245.80]	778.19	22.18%
3	Aged 41–58	Some of the time	All of the time	Often	Not working	*	*	212.37 [190.10 234.64]	787.63	21.24%
	Aged 59+							186.33 l [166.46 206.20]	813.67	18.63%
4	Aged 18–40	Most of the time	All of the time	Often	FT education	Fair	All of the time	245.70 [210.58 280.82]	754.30	24.57%
	Aged 59+							160.15 [142.43 177.87]	839.85	16.02%
5	1st income tertile	All of the time	All of the time	Often	Informal caregiving	Poor	All of the time	63.71 [50.24 77.18]	936.29	6.37%
	2nd income tertile							83.89 [70.02 97.76]	916.11	8.39%
	3rd income tertile							102.88 [90.55 115.20]	897.12	10.23%

Note: Baseline equivalised disposable income 1000(£/m). SIPHER-7 profiles selected to represent the largest difference in WTP between subgroups.
Abbreviations: £/m, pounds per month; CI, confidence interval; EI, equivalent income; WTP, willingness to pay; Y, equivalised disposable income.

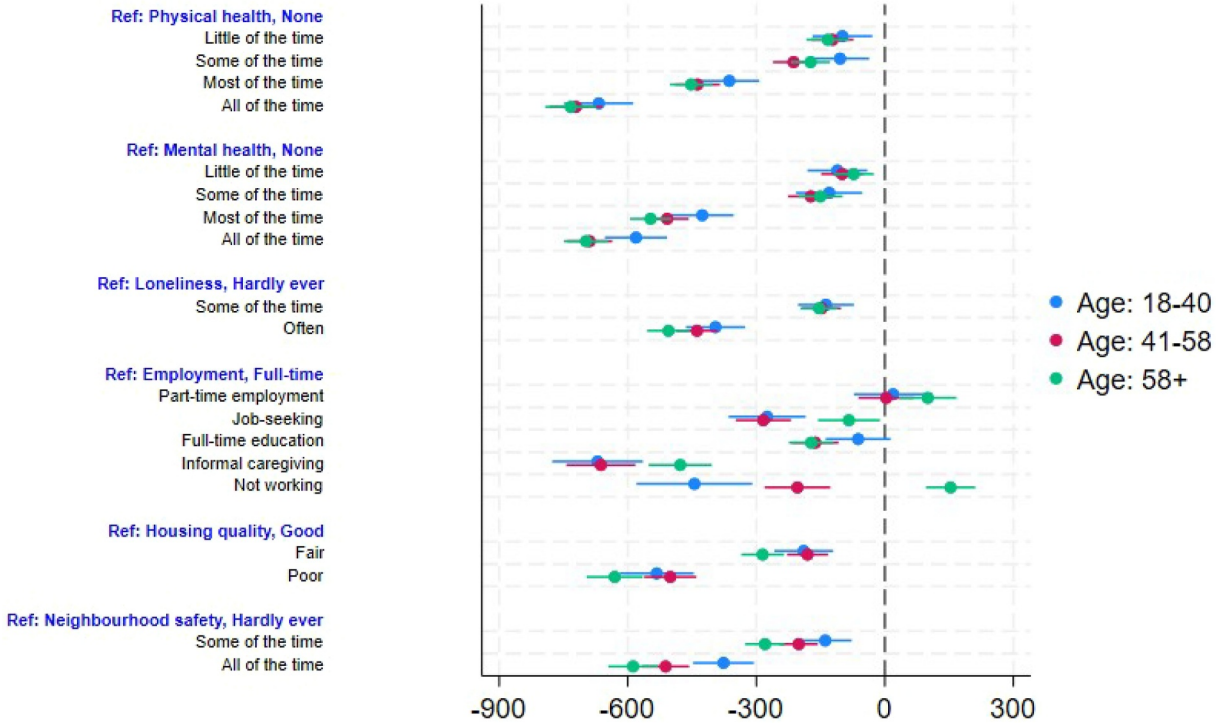

FIGURE 3 WTP for attribute levels by tertile age groups. The plot was based on a mixed logit model allowing random coefficients (non-income). WTP, willingness to pay.

4.2.3 | Subgroup by income

Of the value set sample: 29%, 29% and 42% report household disposable incomes corresponding to the 1st, 2nd and 3rd, and 4th to 10th of the UKHLS income decile groups. Figure 4 illustrates that there is little evidence for preference heterogeneity across these income groups. The largest gaps in equivalent incomes starting from the same hypothetical equivalised disposable income of £1000/month (Table 3) are small and CIs overlap between adjacent income tertile groups, although the difference between the 1st and 3rd income tertile groups have non-overlapping CI, with those in the highest income group having the smallest WTP. The only significant differences are related to having Physical health and Mental health problems "all of the time" and "poor housing" where the 3rd tertile income group has lower WTP.

4.3 | Examining the stability of preferences across survey waves

4.3.1 | Attritors versus stayers: P1 versus C1

To examine the effect of attrition, observations of the Stayers in Wave 1 (P1) and observations of the Attritors in Wave 1 (C1) were modeled separately, and WTP were calculated using the hypothetical equivalised disposable income of £1000/month.

Figure 5 shows that there is no significant difference between the two groups of observations related to the Physical health, Mental health, Loneliness, Housing quality, Neighborhood safety and most of the categories of Employment: the exception is "job-seeking," where the Attritors have larger WTP.

Looking at the six non-income attributes in terms of the WTP to go from the worst level to the best level in each, their ranking is different between P1 and C1. For P1, the WTP is largest for Mental health, followed by Physical health, Neighborhood safety, Housing quality, and Loneliness. For C1, the corresponding ranking is Mental health, followed closely by Physical health, then "informal caregiving," Housing quality, Loneliness and Neighborhood safety.

Amongst those who participated in Wave 1, there is a difference in preferences at the attribute level between the Stayers, who returned for Wave 2 (P1), and the Attritors, who dropped out (C1). The Stayers appear to have a

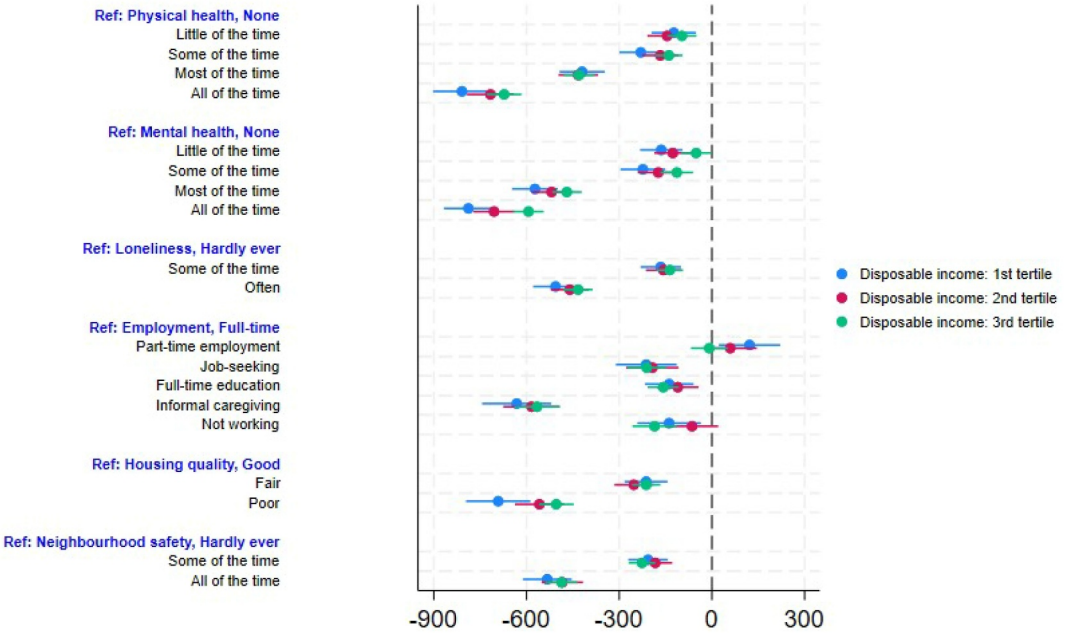

FIGURE 4 WTP for attribute levels by income tertile groups. The plot was based on a mixed logit model allowing random coefficients (non-income). WTP, willingness to pay.

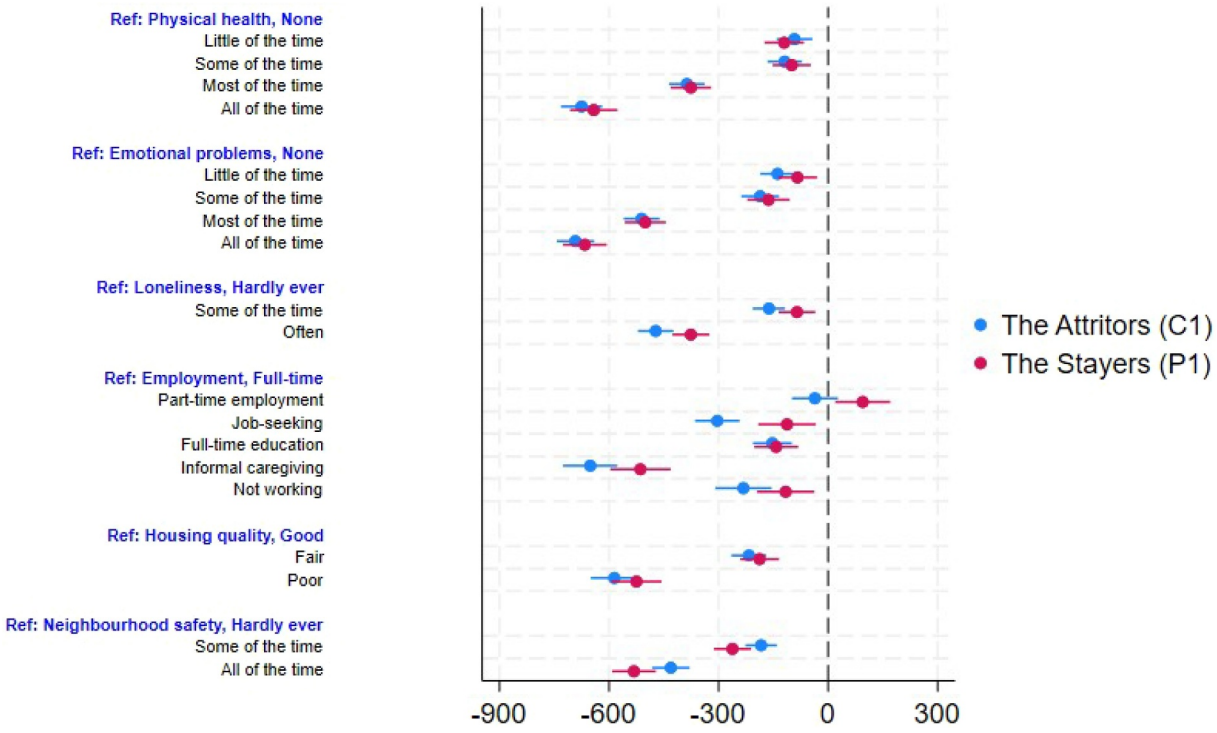

FIGURE 5 WTP for attribute levels by the Attritors (C1) versus the Stayers (P1). The plot was based on a mixed logit model allowing random coefficients (non-income). WTP, willingness to pay.

strong preference for "PT employment" over "FT employment", while the Attritors place more weight on Loneliness and Housing quality and less weight on Neighborhood safety. The biggest gap between the two groups lies in "job-seeking" in the Employment dimension, where the Attritors have a significantly larger WTP than the Stayers.

The first row of Table 4 presents the SIPHER-7 profile made up of the levels where the Attritors have the largest WTP relative to the Stayers. Since the Attritors have smaller WTPs for Neighborhood safety, the best level (where both groups have zero WTP, represented by the asterisk) is used. Stayers have a higher equivalent income than Attritors by up to 6% of equivalised disposable income, which is statistically significant.

4.3.2 | The effect of time: The panel approach and the repeated cross-section approach

In both approaches to examine the effect of time, the observations from the different waves were modeled separately, and WTP were calculated using the hypothetical equivalised disposable income of £1000/month.

Regarding the panel approach, Figure 6 plots the WTP for attribute levels of the balanced panel sample across the two waves (P1 and P2; the Stayers). Overall, the results indicate stable preferences over time. In terms of equivalent income for select SIPHER-7 profiles, WTP for some attributes are larger at Wave 1 than at Wave 2, but not all. Thus, Table 4 reports equivalent incomes for two separate SIPHER-7 profiles: Row 2 is for the profile for which P1 has a larger WTP than P2, while Row 3 is for the profile where P2 has a larger WTP. The profile in which P1 associated with higher WTP results in non-overlapping CIs, indicating significant differences between P1 and P2. Whereas the profiles in which P2 has higher WTP has overlapping CIs amongst the Stayers at Wave 1 and at Wave 2, implying stable preferences over time for this panel sample. It can be concluded that equivalent income based on the autumn 2020 preferences can be statistically significantly higher than that based on the autumn 2021 preferences by almost 16% of equivalised disposable income.

Regarding the repeated cross-section approach, Figure 7 plots the WTPs for the first-timers at Wave 1 (P1 + C1) and the first-timers at Wave 2 (C2). Overall, the results suggest stable preferences at the attribute level with overlapping CIs across all attribute levels. In Table 4, Row 4 is for the profile where the first-timers in Wave 1 have a larger WTP, while Row 5 is for the profile where the first-timers in Wave 2 have a larger WTP. The former profile results in overlapping CIs, whereas the CIs do not overlap in the latter profile. In other words, equivalent income based on the autumn 2020 preferences can be statistically significantly higher than that based on the autumn 2021 preferences by just over 6% of equivalised disposable income. This finding is robust to the application of sampling weights to correct for the different sample composition across the two sets of first-timers (see Appendix 6) and similar to the findings using the panel approach.

TABLE 4 Absolute equivalent income and relative equivalent income for select SIPHER-7 profiles across waves.

	Sets of obs	Physical health	Mental health	Loneliness	Employment	Housing quality	Neighborhood safety	Equivalent income (£/m) [95% CI]	WTP (£/m)	%EI in Y
1	P1	All of the time	Little of the time	Often	Job-seeking	Poor	[a]	175.70 [152.55 198.85]	824.30	17.57%
	C1							113.65 [99.32 127.98]	886.35	11.37%
2	P1	[a]	Most of the time	[a]	Not working	[a]	Some of the time	415.21 [367.88 462.55]	584.79	41.52%
	P2							484.56 [436.40 532.71]	515.44	48.46%
3	P1	Some of the time	Some of the time	Some of the time	Job-seeking	Fair	[a]	643.43 [555.81 731.05]	356.57	64.34%
	P2							485.12 [427.45 542.78]	514.88	48.51%
4	P1 + C1	[a]	Some of the time	[a]	PT employment	[a]	[a]	749.87 [693.08 806.65]	250.13	74.99%
	C2							909.99 [779.81 1040.16]	90.01	91.00%
5	P1 + C1	Some of the time	[a]	Often	[a]	Fair	All of the time	292.96 [270.99 314.93]	707.04	29.30%
	C2							227.06 [194.14 259.99]	772.94	22.71%

Note: SIPHER-7 profiles selected to represent the largest difference in WTP across sets of observations. Baseline equivalised disposable income 1000(£/m). P1: Stayers, C1: Attritors. (P1 + C1): first-timers in Wave 1, C2: first-timers in Wave 2.

Abbreviations: £/m, pounds per month; CI, confidence interval; EI, equivalent income; obs, observations; WTP, willingness to pay; Y, equivalised disposable income.

[a]indicates the reference level of the attribute.

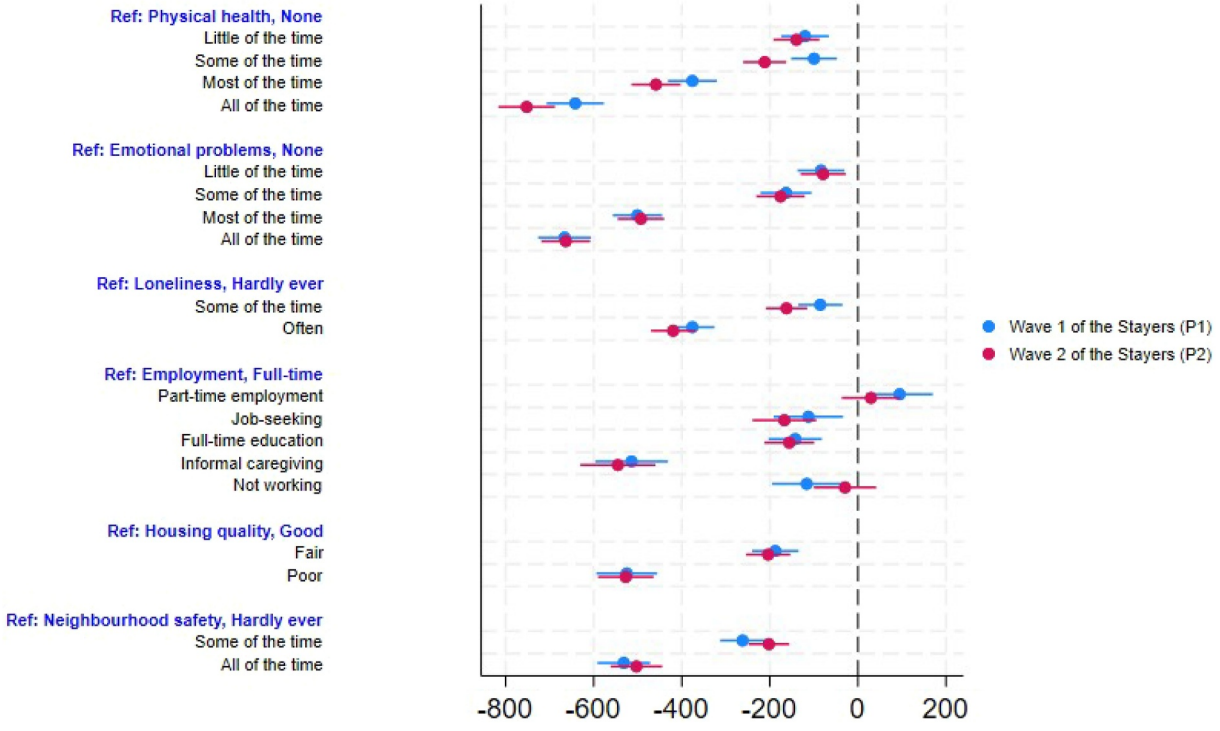

FIGURE 6 WTP for attribute levels using the panel approach: the Stayers in Waves 1 (P1) versus Wave 2 (P2). The plot was based on a mixed logit model allowing random coefficients (non-income). WTP, willingness to pay.

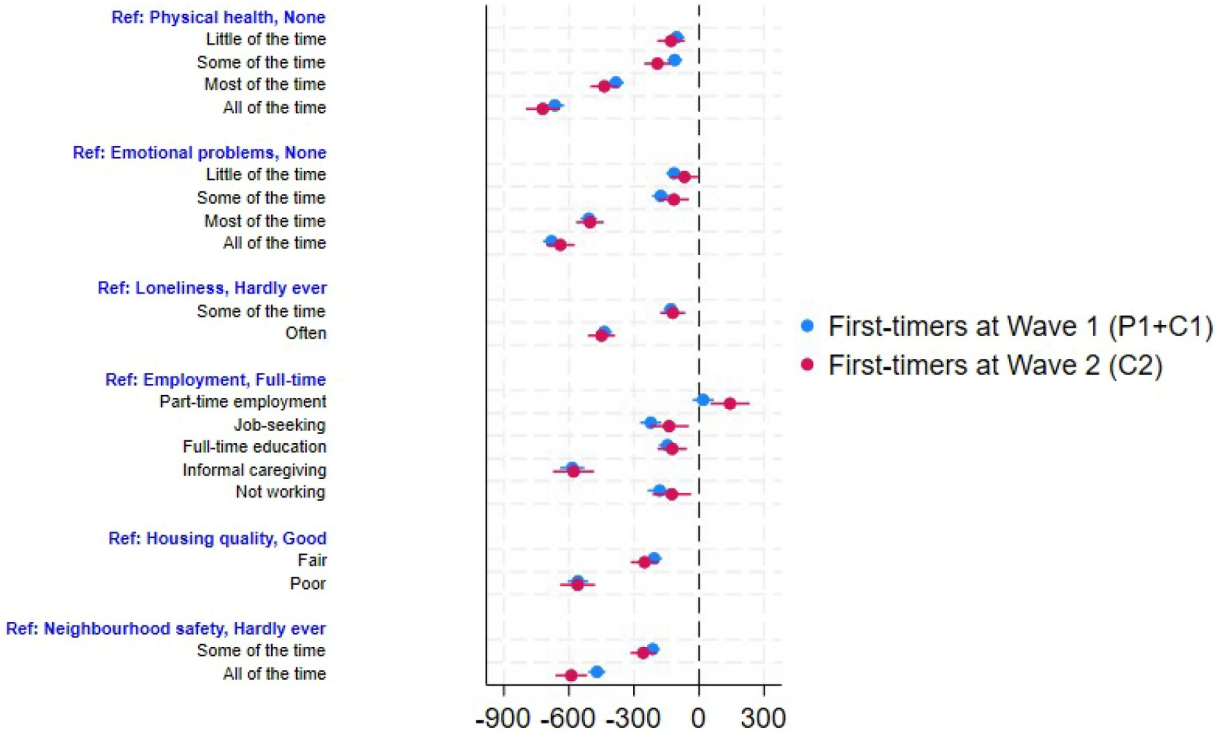

FIGURE 7 WTP for attribute levels using the repeated cross-section approach: first-timers at Wave 1 (P1 + C1) versus first-timers at Wave 2 (P2). The plot was based on a mixed logit model allowing random coefficients (non-income). WTP, willingness to pay.

5 | DISCUSSION AND CONCLUSION

This study has estimated an equivalent income value set for a suite of seven wellbeing indicators, SIPHER-7, using a DCE conducted online across two waves, each with a representative sample of more than 3300 respondents from the UK general public. The econometric models generated significant coefficients with the expected signs, where relevant, which were then used to predict equivalent income for SIPHER-7 profiles.

This is the first large-scale empirical study that we are aware of to estimate an equivalent income value set using stated preference data of a nationally representative general public sample. Since the wellbeing indicators used in SIPHER-7 can be linked back to the UKHLS, the equivalent income value set reported here allows researchers to calculate a preference-based single index of wellbeing for respondents in the survey. Our findings indicate that, of the six non-income wellbeing indicators used, Physical health and Mental health are the most important, followed by Employment, Housing quality, Neighborhood safety and Loneliness.

Existing computations of equivalent income have mostly relied on subjective wellbeing (SWB) or satisfaction regressions using national surveys. However, WTP values elicited through happiness functions often appear unrealistically high, and there is growing evidence that coefficients from happiness regressions might not be suitable to calculate reliable marginal rates of substitution (Benjamin et al., 2014).

The design of our stated preference survey involved a methodological challenge regarding the presentation of the monetary attribute in the DCE. In contrast to most DCE contingent valuation studies that elicit WTP as the variable of interest, this study aimed to obtain equivalent income, which is the amount of household disposable income left after subtracting the WTP to achieve the best levels of non-income dimensions. This resulted in two complications: first, the income levels in the DCE had to be amounts that respondents could relate to. We used relative income levels to generate a single DCE design, but presented these in terms of absolute income levels based on the income range each respondent self-reported their household income to be in. Second, since equivalent income is an individual level measure, household income in the DCE needed to be corrected for household size, and this was done by equivalising the household income level shown in the DCE at the analysis stage using the household size information for each respondent.

Regarding group-level preference heterogeneity, value sets in health outcome measures, such as those for the EQ-5D, have traditionally not explored heterogeneous preference by population subgroups. This is likely to be because, while preference heterogeneity in health is likely to exist, resource allocation in publicly funded health care has typically assumed that there is one population preference based on which priority setting should be carried out. By reporting subgroup value sets, we do not intend to make a normative argument over the use or otherwise of subgroup value sets in policy evaluation. But there may be instances with policy interests in examining the preferences of relevant subpopulations: for example, policymakers may be interested in how a particular intervention might be perceived by the target population. If so, the study reports results by gender groups, by age groups and by income groups, all of which resulted in significant differences depending on SIPHER-7 profiles. In terms of equivalent income, these differences were found to be at the most under 12% (and significant) of baseline equivalised disposable income between the gender groups and for age groups, and at the most under 5% (and significant) across income groups.

With policy relevance in mind, the analysis defined subgroups in terms of one observable characteristic at a time. The results from the mixed logit estimations used to control for individual-level preference heterogeneity were largely similar to those obtained from pooled conditional logit estimations with robust errors (for individual-level clustering) - the latter results are in Appendix 5. Furthermore, if the objective were to analyze individual-level preference heterogeneity itself alongside observable group-level heterogeneity, more complex data-driven modeling approaches are available, such as the use of latent class analysis with observable characteristics in the class-membership, or interacting observable characteristics with random parameters.

The study also explored preferences over survey waves, where the second wave, conducted roughly 12 months later, not only re-invited the respondents to the first wave, but also a fresh sample recruited in the same way from the same population. This design allowed us to distinguish between the effect of attrition and the effect of time. The analysis found the presence of both. Regarding the effect of attrition, the Attritors and the Stayers have clearly distinct preferences. The effect of time was explored in two ways. The first, panel, approach found no significant change in preference between the two waves. However, the second, repeated cross-section, approach found that first-time respondents in autumn 2020 and first-time respondents in autumn 2021 have significantly different preferences depending on SIPHER-7 profiles, amounting to up to 7% of baseline equivalised disposable income (which was robust to correcting for the different sample composition across the two sets of first-time respondents). This suggests that the panel approach

may overestimate the extent of the stability of preferences over time. While there are some studies that use the panel approach to examine the stability of DCE preferences (e.g., San Miguel et al., 2002), we are not aware of any DCE studies that contrast the effect of attrition and the effect of time.

The study has some shortcomings. The suite of wellbeing indicators, SIPHER-7, was given to the study. In addition, the design and analysis assume scale and preference homogeneity and do not model preferences at the individual level. Both waves of the survey were conducted online with no insight into respondents' comprehension. In addition, there are limitations related to recruiting respondents through an internet panel, such as the representativeness of the sample with respect to income or internet use.

To conclude, the study provides a population value set in the form of equivalent income, for a suite of wellbeing indicators called SIPHER-7, based on preferences of the UK general public. Since the dimensions of SIPHER-7 are linked to the UKHLS survey, the results can be used to calculate equivalent income for any UKHLS respondent. Furthermore, the outputs have the potential to equip policy-makers with a preference-based and interpersonally comparable metric of likely public responses to changes in policies.

ACKNOWLEDGMENTS

This work was carried out as part of the SIPHER Consortium (in Systems Science in Public Health and Health Economics Research). An earlier version of this paper was presented at the Health Economists' Study Group meeting (Cambridge, June 2021). We are grateful for comments and suggestions from Duncan Chambers, Dan Chedgzoy, Koen Decancq, Lucy Gavens, Mary Gogarty, Daniel Gray, Arne Risa Hole, Ceri Hughes, Petra Meier, Rebecca McDonald, Robin Purshouse, Ellen Stewart, Tara Wickramasekera, Hui Zhang and the two anonymous Referees from Health Economics. We would also like to thank all survey participants who took part. The usual disclaimers apply. The study was approved by the ethics committee at the Department of Economics, the University of Sheffield. An Ta was funded by the Department of Economics, the University of Sheffield. SIPHER Consortium funded the fieldwork of this study. SIPHER Consortium is supported by the UK Prevention Research Partnership (MR/S037578/2), which is funded by the British Heart Foundation, Cancer Research UK, Chief Scientist Office of the Scottish Government Health and Social Care Directorates, Engineering and Physical Sciences Research Council, Economic and Social Research Council, Health and Social Care Research and Development Division (Welsh Government), Medical Research Council, National Institute for Health Research, Natural Environment Research Council, Public Health Agency (Northern Ireland), The Health Foundation and Wellcome. For the purpose of open access, the authors have applied a Creative Commons Attribution (CC BY) license to any Author Accepted Manuscript version arising.

CONFLICT OF INTEREST STATEMENT

We declare that there are no conflicts of interest and no competing interests.

DATA AVAILABILITY STATEMENT

The authors agree to depositing all data including the choice experiment design at the University of Sheffield's Online Research Data (ORDA).

ORCID

Aki Tsuchiya https://orcid.org/0000-0003-4245-5399

REFERENCES

Abasolo, I., Sandelind, C., Schokkaert, E., Stevens, K., & Tsuchiya, A. (2018). *Operationalising equivalent consumption through stated preferences*. CwiPP Working Paper No. 13. Centre for Wellbeing in Public Policy, University of Sheffield.

Benjamin, D., Heffetz, O., Kimball, M., & Rees-Jones, A. (2014). Can marginal rates of substitution be inferred from happiness data? Evidence from residency choices. *The American Economic Review*, *104*(11), 3498–3528. https://doi.org/10.1257/aer.104.11.3498

Capéau, B., Cherchye, L., Decancq, K., Decoster, A., De Rock, B., Maniquet, F., Nys, A., Périlleux, G., Ramaekers, E., Rongé, Z., & Schokkaert, E. (2020). *Well-being in Belgium*. Springer International Publishing.

ChoiceMetrics Pty Ltd. (2018). Ngene user manual and reference guide. Accessed 03 February 2020.

Decancq, K., Fleurbaey, M., & Schokkaert, E. (2015). Happiness, equivalent incomes and respect for individual preferences. *Economica*, *82*(s1), 1082–1106. https://doi.org/10.1111/ecca.12152

Decancq, K., & Schokkaert, E. (2016). Beyond GDP: Using equivalent incomes to measure well-being in europe. *Social Indicators Research*, *126*(1), 21–55. https://doi.org/10.1007/s11205-015-0885-x

Dolan, P., & Roberts, J. (2002). To what extent can we explain time trade-off values from other information about respondents? *Social Science & Medicine*, *54*(6), 919–929. https://doi.org/10.1016/s0277-9536(01)00066-1

Fisher, P. (2019). Does repeated measurement improve income data quality? *Oxford Bulletin of Economics & Statistics*, *81*(5), 989–1011. https://doi.org/10.1111/obes.12296

Fleurbaey, M. (2005). Health, wealth, and fairness. *Journal of Public Economic Theory*, *7*(2), 253–284. https://doi.org/10.1111/j.1467-9779.2005.00203.x

Fleurbaey, M. (2006). Health, equity and social welfare. *Annales d'Economie et Statistique*(83/84), 21–59. https://doi.org/10.2307/20079162

Fleurbaey, M. (2009). Beyond GDP: The quest for a measure of social welfare. *Journal of Economic Literature*, *47*(4), 1029–1075. https://doi.org/10.1257/jel.47.4.1029

Fleurbaey, M., & Blanchet, D. (2013). *Beyond GDP: Measuring welfare and assessing sustainability*. Oxford University Press.

Fleurbaey, M., & Gaulier, G. (2009). International comparisons of living standards by equivalent incomes. *The Scandinavian Journal of Economics*, *111*(3), 597–624. https://doi.org/10.1111/j.1467-9442.2009.01578.x

Fleurbaey, M., Luchini, S., Muller, C., & Schokkaert, E. (2013). Equivalent income and fair evaluation of health care. *Health Economics*, *22*(6), 711–729. https://doi.org/10.1002/hec.2859

Fleurbaey, M., & Schokkaert, E. (2011). Equity in health and health care. In M. V. Pauly, T. G. McGuilre, & P. P. Barros (Eds.), *Handbook of health economics volume two*.

Hole, A. R. (2007). Fitting mixed logit models by using maximum simulated likelihood. *STATA Journal*, *7*(3), 388–401. https://doi.org/10.1177/1536867x0700700306

OECD. (2008). Growing unequal? Income distribution and poverty in OECD countries.

OECD. (2011). Divided we stand – why inequality keeps rising.

Quattrociocchi, J. (2018). Group income and individual preferences for redistribution. *Canadian Journal of Economics/Revue canadienne d'économique*, *51*(4), 1386–1418. https://doi.org/10.1111/caje.12358

San Miguel, F., Ryan, M., & Scott, A. (2002). Are preferences stable? The case of health care. *Journal of Economic Behavior & Organization*, *48*(1), 1–14. https://doi.org/10.1016/s0167-2681(01)00220-7

Stiglitz, J. E., Sen, A., & Fitoussi, J. P. (2009). Report by the Commission on measurement of economic performance and social progress. Technical report.

Tangian, A. S. (2005). *A composite indicator of working conditions in the EU-15 for policy monitoring and analytical purposes*. WSI Diskussionspapier 135 (p. 77). Hans Böckler Stiftung, Düsseldorf.

Tsuchiya, A., & Wu, C., & on behalf of the SIPHER Consortium. (2021). SIPHER-7: A seven-indicator outcome measure to capture wellbeing for economic evaluation. Working paper, SIPHER Research Paper Series 1. Retrieved from https://sipher.ac.uk/wp-content/uploads/2021/10/Sipher-7-report.pdf

University of Essex, Institute for Social and Economic Research. (2020). *Understanding society: Waves 1-10, 2009-2019 and harmonised BHPS: Waves 1-18, 1991-2009. [data collection]* (13th ed.). UK Data Service. SN: 6614.

Ura, J. D., & Ellis, C. R. (2008). Income, preferences, and the dynamics of policy responsiveness. *PS: Political Science & Politics*, *41*(4), 785–794. https://doi.org/10.1017/s104909650808102x

Van Landeghem, B. (2014). A test based on panel refreshments for panel conditioning in stated utility measures. *Economics Letters*, *124*(2), 236–238. ISSN 0165-1765. https://doi.org/10.1016/j.econlet.2014.05.024

Van Landeghem, B. (2019). Stable traits but unstable measures? Identifying panel effects in self-reflective survey questions. *Journal of Economic Psychology*, *72*, 83–95. https://doi.org/10.1016/j.joep.2019.02.006

World Health Organization. (2014). Health in all policies: Helsinki statement. Framework for country action.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Ta, A., Van Landeghem, B., & Tsuchiya, A. (2024). Eliciting public preferences across health and wellbeing dimensions: An equivalent income value set for SIPHER-7. *Health Economics*, 1–19. https://doi.org/10.1002/hec.4890