Pre-Analysis Plan for the ANR JCJC project 'Improvements on valuing gains in health status with stated preferences' (ANR-20-CE36-0010-01)

1 Introduction

The project ANR-20-CE36-0010-01 (VHEALTH hereafter) aims at providing new tools to assess monetary values for improved health and longevity. The most common practice to approximate the monetary value of a gain in health or longevity (or both) is to combine preference-based measures of the severity and duration of illness, often expressed in quality-adjusted life years (QALYs), with monetary estimates of the VSL (e.g. Mason et al., 2009). The QALY metric is widely used in medical and health economics. It assumes that health is the sum of the time spent in each health state weighted by an index of health quality, where a weight of one corresponds to full health and a weight of zero to a health state as bad as dead. The guidelines for regulatory impact analysis of the U.S. Department of Health and Human Services (2016) recommend that an "approach [for valuing morbidity] is to calculate a constant value based on expected QALYs [...] by dividing the VSL by future QALYs". Dividing VSL by future QALYs implicitly assumes that the tradeoff between wealth and the probability of surviving the current period is proportional to future QALYs. However, this assumption is inconsistent with empirical evidence on how the VSL varies with age (Aldy and Viscusi, 2007; Krupnick, 2007). The approach implies that the value for improved health or longevity is constant, which violates standard assumptions of mortality risk valuation (e.g., VSL varying with health and longevity; Hammitt, 2013).

In previous work (Herrera-Araujo et. al, 2020) we explore the validity of the widely implemented practice of dividing VSL by future QALYs to approximate the monetary value of a health improvement. Our theoretical setting allows us to derive empirical tractable upper and lower bounds for the value for improved health and longevity. We find that dividing VSL by future QALYs corresponds to an upper bound of the WTP for improving health, which implies that dividing VSL by future QALYs overstates the monetary value of health gain. We also identify a lower bound not previously used in policy applications but appealing given its conservative nature. Our research aims at empirically identifying the upper and lower bounds for the value of improved health and longevity along with VSL estimates in France

The project's second objective aims at providing empirical support for an innovative method for estimating the VSL through SP: the use of non-marginal risk reductions. Most papers ask

respondents to value small risk reductions to elicit VSL. These have the advantage that under conventional theoretical assumptions, WTP should be nearly proportional to the proposed risk reduction for small risk reductions.

For example, if WTP for a 1 in 100,000 change in the annual risk of premature death equals 30 euros, then the WTP for a 2 in 100,000 change should be close to 60 euros. This result provides a useful scope test: responses should be consistent with the hypothesis that WTP is nearly proportional to the risk reduction. Passing a scope test is considered a necessary condition for an SP study to be of good quality. In practice, stated-preference studies pervasively suffer from a lack of scope sensitivity (Hammitt and Graham, 1999). This is called "scope insensitivity". There are at least two reasons for this. First, scope insensitivity may result from the diminishing rate at which individual's trade income for risk reduction as the risk change increases (i.e., the curvature of the utility function; Corso et al. 2001); and second, it may result from respondents' lack of understanding of the "good" being valued.

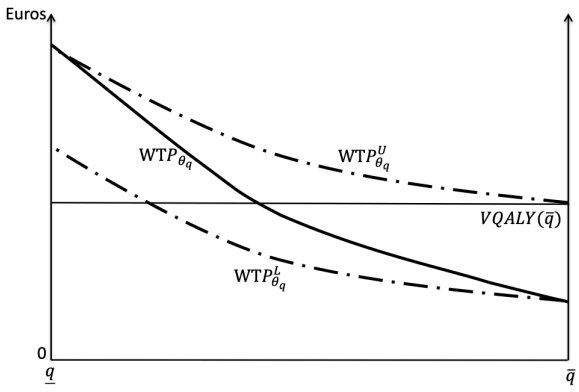
In designing VSL elicitation questions, there is a trade-off between distortions due to the curvature of the utility function and distortions due to respondents' limited comprehension of the risk reduction. For small risk reductions (say, a 1 in 100000 change in the annual risk of premature death), the curvature of the utility function -the rate of change in the rate at which individual's trade income for risk reductions- is unimportant, but the respondents' cognitive cost of evaluating such small changes is high. For larger risk reductions (e.g., a 1 in 100 change in the annual risk of premature death), the curvature becomes more relevant and respondents should have a better understanding of the size of the risk change they are valuing.

2 Research Design

2.1 Hypotheses

As markets for mortality risk reductions do not exist, researchers have used non-market valuation techniques such as revealed preference (RP) and SP methods. RP methods infer individuals' preferences from their behavior in contexts that affect their mortality risks, most often by analyzing compensating wage differentials for occupational fatality risk (Viscusi and Aldy, 2003; Herrera-Araujo and Rochaix, 2020). SP methods are more controversial than RP methods, yet they remain one of the principal methods for estimating the rate of substitution between wealth and small changes in health risks. One reason is that SP methods are more flexible than RP methods since researchers can specify the characteristics of the intervention to be valued and the population whose preferences are elicited. Specific to our context, using SP methods allows us to elicit key ingredients for each of the two proposed studies.

In order to guide the empirical strategy and its interpretation we follow the theoretical model developed in Herrera-Araujo et al. (2020). The model studies the properties of WTP for improved health and extended longevity within a framework that relies on minimal assumptions about an agent's life-time utility function and the interactions between their wealth, health and



longevity prospects. We assume that preferences for health and longevity can be summarized using a scalar-valued index (such as QALYs) and utility gained from wealth (or consumption potential) is separable from this index. In this section, we refer to the scalar-valued index as 'quality of life' and use it interchangeably with 'health and longevity'. The empirical predictions of the model can be summarized with the following figure:

The bold line represents the WTP for a marginal improvement in quality of life. The upper and lower bounds are depicted by the dashed lines above and below the bold line. The first theoretical prediction is that both the WTP and its bounds are decreasing with better baseline quality of life. This implies that individuals with a lower baseline quality of life, either due to shorter longevity (i.e., older individuals) or worse health related quality of life (i.e., individuals suffering from cancer), have a higher WTP for a marginal improvement in quality of life than better-off individuals. This is consistent with the so-called dead-anyway effect in the VSL literature (Pratt and Zeckhauser, 1996) and reflects the low private opportunity cost of spending for individuals near death. From a policy perspective, this finding advocates against the use of a one-size-fits-all value of improved quality of life (often obtained by dividing a constant VSL by future QALYs). The second theoretical prediction is that under reasonable assumptions, the WTP will differ from its bounds. In turn, this allows us to provide a new empirical test for the descriptive validity of the QALY metric: differing values for the upper and lower bounds provide evidence against the validity of the QALY metric. Our model provides expressions that are empirically tractable to elicit the upper and lower bounds.

The empirical identification of the upper bound on WTP for improved quality of life depends on two ingredients: quality of life and the VSL. To empirically identify the lower bound an additional ingredient is needed: the results from a generalized standard gamble. The standard gamble measures the utility of a certain health state by comparing it with two other health states, say death and perfect health. This gamble elicits the probability of immediate death which would make an individual indifferent between a sure outcome in which they would live for a period of life in a health state corresponding to a bad quality of life and a lottery between immediate death and full quality of life over the same period. To elicit the three ingredients within a single study requires the use of a SP survey. The project develops an original survey design that captures the three required elements to identify the upper and lower bounds.

2.2 Framework - Survey structure

The type of information provided to respondents, the payment vehicle, the explanation of the risk reduction, the choice of stimulus and levels of characteristics were all carefully considered in the design of the survey instrument. We built on an online survey that we had previously implemented in France (Hammitt and Herrera-Araujo, 2018) and the US (Hammitt and Haninger, 2010). That survey outlined a hypothetical mortality risk associated with pesticide residues in food and proposed risk reductions that are obtained by purchasing an alternative food produced following a hypothetical, "pesticide safety system" supervised by public authorities.

The survey instrument is divided into the following sections: consent, demographics for quota-setting purposes, practice questions, WTP question, standard gamble, financial risk aversion and general follow-up questions and further socio-economic questions. In the consent statement, respondents are informed about the objectives of the survey, the approximate time it will take to complete the survey (we aim at 25 minutes), their data rights, and the compensation awarded by the recruiting platform for participating. Next, respondents will be asked about their demographics and their current health status. Several measures are used here to measure health. Current health is elicited using either a visual analog scale (VAS) on which 100 corresponds to full health and 0 to a health state equivalent to death, or the EQ-5D health state classification system (EuroQol Group, 1990). The EQ-5D is a utility instrument used to quantify the health-related quality of life (HRQL) associated with an individual's current health or a hypothetical health state. It classifies health states using five dimensions (mobility, self-care, usual activities, pain/discomfort, anxiety/depression), each of which can take any of three levels (no, moderate, or severe problems) (or five levels in a newer version). HRQL is an index normalized to a value of one for full health and zero for health states an individual judges indifferent to being dead (negative values are permitted). It can be estimated from the EQ-5D health-state description using a scoring rule (Andrade et al., 2020).

To elicit the upper and lower bounds on the WTP for improved health (Herrera-Aurajo et al., 2020), we require two additional elements besides health status. First, a measure of each individual's WTP for a small mortality risk reduction (i.e., VSL), which requires a valuation scenario. Second, a measure of the individual's indifference between a certain health state and a lottery that would either improve quality of life to its maximum or cause immediate death. The common indifference are the odds that make the individual indifferent, which can be elicited using a standard gamble. The results of the standard gamble are then combined with the WTP results to determine both upper and lower bounds.

The upper bound can be directly obtained using a combination of the valuation question with each individual's current quality of life estimate. There are several challenges with the design of the valuation questions that have to be addressed, in particular, making sure that respondents understand the good that is being valued. We deploy state-of-the-art risk communication tools (Spiegelhalter, 2017) to improve respondents' understanding. The valuation section follows a double-bounded binary-choice format (Hanemann et al., 1991). In this type of format the

respondent is presented with two bids (amounts to be paid to partake in the treatment). The initial bid (the incremental cost of the intervention) is randomly assigned; the follow-up bid is twice the initial bid for respondents who indicated in the initial question that they would choose the treatment, and half the initial bid for the other respondents.

We draw from experiences with the so-called 'chained approach' (e.g. Pinto-Prades et al., 2009; Baker et al., 2010; Pennington et al., 2015) for the estimation of the lower bound. This approach breaks down the valuation process into a series of steps involving responses to WTP and standard gamble questions (Carthy et al., 1999). Analogous to the chained approach, the first stage uses the monetary estimate derived from the estimation of the upper bound. In the second stage, respondents will be presented with a standard gamble. The monetary value from the first stage will then be combined with the standard gamble results from the second stage to obtain the lower bound.

As this is a novel survey instrument, we tested it extensively using ten one-to-one interviews, and two-focus groups making sure that the survey instrument will be well understood by the respondents. Panel members are to be recruited by random sampling to closely match the French adult population in terms of age, gender, socio-professional categories, regions and population variables. Respondents are compensated in a way consistent with the platform's standard method (points for online purchases). We aim at gathering 5,150 completed surveys.

2.3 Intervention - Experimental design

2.3.1 WTP valuation

Within the WTP task, we ask respondents to consider a status-quo situation described by five attributes: (1) risk of developing cancer and dying from it; (2) risk of developing cancer and recovering from it; (3) quality-of-life conditional on having cancer; (4) duration of cancer spell; and (5) cost for buying regular food products. Respondents are asked to compare it with a scenario where, at a higher cost, the 'risk of developing cancer and dying from it' and/or the 'risk of developing cancer and recovering from it' are reduced via the purchase of /safety-certified food products.

We consider three cancer risk horizons: 5 years, 10 years and 20 years. Each respondent is randomly allocated at the beginning of the survey to one of the three horizons. Once allocated, the horizon remains fixed throughout the survey. The levels for both risk attributes are adapted to match the horizons.[1] We communicate risks in terms of 'X in 1000', where X is the number of cases per 1,000 individuals, using a visual aid in the form of grids. Each grid contains 1,000 - X grey cells and X color-blind friendly marked cells. Prior to the elicitation, respondents receive probability training and feedback using these grids for risk communication.

We focus on 'cancer' without explicitly mentioning a specific type of cancer, nor the affected organ. We describe the severity of the disease by a combination of duration of the effects and the

[1]The 5-year/10-year risks differ by a factor of 0.5, while the 10-year/20-year by a factor of 2.5. These factors are based on average life-time cancer risks using data from European Cancer Information System.

corresponding health-related quality (HRQL). The HRQL is represented with a visual analogue scale using QALY weights as inputs (Andradade et al, 2020). Half of the respondents randomly receive either a 'bad' quality-of-life or a 'very bad' quality of life.[2,3] Following the same pattern, half of the respondents are randomly assigned to a short duration (2 years) or a long duration (4 years) of health effects. The quality-of-life with cancer (both severity and duration) is the same across alternatives within respondents, but varies across respondents.

The valuation exercise of the questionnaire is comprised of three double-bounded dichotomous-choice tasks. In each task respondents choose between the safer-but-costlier option, or the status-quo (i.e., is riskier-but-cheaper). If the respondent opts for the safer option, the initial bid is doubled. If the respondent opts for the status quo, the initial bid is halved. An additional third elicitation question is asked only if the respondent answered YES-YES or NO-NO. In the YES-YES case, the initial bid is multiplied by 6, while in the NO-NO case, the respondent is asked whether they are willing to pay one-tenth of the initial bid.[4] The alternatives are presented using a graphical format where the status quo option and the safer option are compared attribute-by-attribute.

We ask respondents to think of these three valuation tasks as distinct and independent scenarios. We emphasise that the risk reduction offered in each valuation task is experienced only by the respondent and not by other household members. The risk reduction is linked to pesticides and other chemical substances in food. We describe that the safer food is certified by an official 'food safety program', and by focusing on foods we emphasize that the risk reduction is a private good.

For the experimental design, we create 30 blocks, with different versions of the three double-bounded questions, by randomly sampling the levels of the attributes from the full factorial design.[5] We impose several restrictions: (1) we focus on 'not-too-lethal' cancers, i.e. we only allow for combinations where 'the risk of developing cancer and recovering from it' is larger than 'the risk of developing cancer and dying from it'; (2) respondents are exposed to one set of baseline values; (3) the cost attribute of the status-quo is set to zero, while it is always positive for the safer option; (4) each respondent sees a specific combination of levels (i.e. a specific choice card) only once; (5) the initial bid levels for a choice card differ across scenarios; (6) the risk reduction attributes are never simultaneously equal to zero, i.e. the safer option always

[2]We use the EQ5D-5L version of the questionnaire. The HRQL produced by French EQ5D-5L (Andradade et. al, 2020) normalizes HRQL of a diseases equivalent to death to 0. The HRQL scale ranges from -0.52 to 1, implying that there are worst conditions than death. We multiply the HRQL by 100 for use in our visual analogue scale. We limit the HRQL to the unit interval ($\times 100$). That is, we present to respondents an HRQL scale ranging for a disease equivalent to death, to full health. The rescaled HRQL values for the 'bad' state (i.e., EQ5D-5L 43344) and that for the "very bad" state (i.e., EQ5D-5L 43354) are 27 and 9, respectively.

[3]We make sure that cancer never implies an improvement in health for a respondent. For this, we set each attribute in the cancer EQ5D-5L profile presented to the respondent to be the maximum (i.e., worst) of the level in the assigned cancer EQ5D-5L and the level corresponding to the target's current health. This implies that for some respondents the "bad" or "very bad" health states will differ from 27 or 9. We plan to drop respondents for which our algorithm produces HRQL lower than 0.

[4]In the debriefing section we ask respondents that were not willing to pay 0.1X the initial amount about their reasons for not buying the risk reduction. In particular, we ask respondents if they did not believe the scenario.

[5]We use the same 30 blocks for each cancer risk horizon. To adapt the blocks to the 5-year risk horizon, we multiply the 10-year risk by 0.5, and for the 20-year risk horizon, we multiply the 10-year risk by 2.5.

entails a reduction in risk.

Table 1: Attributes and levels

Attributes	**Levels**
Regular food	
Risk of developing cancer and dying from it	25 in 1000, 50 in 1000
Risk of developing cancer and recovering from it	25 in 1000, 50 in 1000, 100 in 1000
Certified food	
Risk of developing cancer and dying from it (in 1000)	Decreasing the baseline by: 0, 4, 6, 8, 14
Risk of developing cancer and recovering from it (in 1000)	Decreasing the baseline by: 0, 3, 5, 7, 13
Cost per month	100 euros, 200 euros, 280 euros, 400 euros
Common to regular and certified	
Cancer risk horizon	5 years, 10 years, 20 years
EQ5D-5L conditional on having cancer (HRQL)	43344 , 43354
Associated rescaled HRQL for EQ5D-5L 43344	27
Associated rescaled HRQL for EQ5D-5L 43354	9
Duration conditional on having cancer	2 years, 4 years

2.3.2 Standard gamble

The standard gamble asks respondents to assume that they are diagnosed with cancer and two alternative treatments are available: a low-risk (or no risk) treatment that allows them to (nearly) avoid death from cancer, but with which their health status remains compromised for a specified duration; or a high-risk experimental treatment that is highly effective in some people but not effective in others. We assign half of the respondents to a "no-risk" vs. "high-risk" treatment, while the other half are assigned to a "low-risk" vs. "high-risk" treatment. The "low-risk" treatment randomly assigns between a risk of 1 in 1000 or 2 in 1000 of dying, even if taking the treatment.

For each respondent, the duration of effects and the corresponding HRQL are the same as for the WTP questions. We use the conditional mortality probability generated by the first valuation question, after applying the risk reduction. To back out this probability, we proceed as follows: let q denote cancer incidence and p denote the conditional cancer mortality; then the status-quo 'risk of developing and dying from it' is computed as $q \times p$. Similarly, the 'risk of developing and recovering from it' equals $q \times (1 - p)$. Let us denote the risk reductions on incidence and conditional mortality by (θ_q and θ_p, respectively. The reduced risks are accordingly computed as $(q - \theta_q) \times (p - \theta_p)$ and $(q - \theta_q) \times (1 - (p - \theta_p))$. We back out q, p, θ_q and θ_p from the first valuation question of each of the 30 blocks, and use $(p - \theta_p)$ as the baseline for the standard gamble.

As with the WTP question, the standard gamble is comprised of one triple-bounded dichotomous-choice task.[6]. If the respondent opts for the safe treatment, then the initial mortality probability is halved. If the respondent opts for the risky treatment, the initial probability is multiplied by 1.5. If the respondent opts for the safe (risky) alternative, the follow-up question prompts them to select a risk of dying resulting from the risky treatment that would lead them to choose the

[6] An additional third trade-off question is only asked if the respondent chose Safe-Safe or Risky-Risky.

risky (safe) treatment instead. The range of choices for respondents opting always for the safe treatment is $[0, (p-\theta_p)/2]$, while the range for respondents opting always for the risky treatment is $[(p-\theta_p) \times 1.5, 1]$.

2.3.3 Financial risk aversion

Following the WTP question, the standard gamble and some debriefing questions, respondents receive, at random, one of three alternative sets of questions about hypothetical financial risks.

The first set of questions ask each respondent to assume that he/she is the only income earner in the household and has a good job guaranteed to pay the same annual income for life (Barsky et al., 1997). The respondent is offered the opportunity to take a new and equally good job with equal chances that will either double his or her income or cut it by a third. If the respondent accepts the new job, a follow-up question increases the income loss to one-half. On the contrary, if the respondent declines the new job, then a follow-up question reduces the income loss to one-fifth. Responses to the two questions allow us to categorize respondents into four ordered classes of relative risk aversion.

Following Hammitt & Haninger (2010), a second set of questions instructs respondents to decide about two job offers. Specifically, they can choose between a salary job, which pays an annual salary that increases modestly each year, and a bonus job, which pays a smaller salary but offers with some probability a large annual bonus.

Finally, we adapt a version of the Eckel and Grossman risk elicitation task (2002, 2008) which consists of asking subjects to choose their preferred gamble from among six possible choices. All gambles are coin flips, i.e. they offer a 50/50 chance of a low or a high payoff. Payoffs range from a sure payoff (tail and head result in the same payoff) to increasingly higher-risk/higher-reward gambles. Table 2 reports the gambles used for this question.

Table 2: Payoffs and CRRA range: Eckel and Grossman lottery

Gamble	Low payoff	High payoff	Expected return	Std.	CRRA range
G1	50	50	50	0	2.49<r
G2	40	70	55	15	0.84<r<2.49
G3	30	90	60	30	0.5<r<0.84
G4	20	110	65	45	0.5<r<0.63
G5	10	150	80	70	0.39<r<0.63
G6	0	200	100	100	0<r<0.39

3 Data

3.1 Data collection and processing

We plan to administer the survey to a random sample of the French KANTAR Online Panel. Panel members are recruited by random sampling using email and closely match the French adult population on age, gender, socio-economic and geographical variables. Respondents are

compensated with points that can be used to purchase items online. We aim to gather data in three waves between October and December 2023, with 5150 surveys completed. The first wave and second will contain responses from 150 respondents, and 500 respondents. Both will serve as pilots. The third wave will contain 1500 completed surveys, and the fourth wave should contain the remaining completed questionnaires.

In additiona, using respondents self-reported zip-code, we plan to collect and match commune level demographic information with respondents' zip-code of residence.

3.2 Variation from the intended sample size

KANTAR's panel is large enough to ensure 5150 completed surveys. The key challenge is to ensure respondents' attention through the survey. We prepared a set of attention checks within the survey. The first attention check asks respondents to select the category "not much" amongst a set of similarly phrased options. A respondent that reads the question should be able to pick up the correct option quite easily. If the respondents answer wrongly this questions, they are filtered out of the survey.

Attention can be proxied by the time spent on the survey. The pilot will provide survey length information for identifying and dropping speedsters. If a respondent completes a survey in less than 0.4 X median time from reported in the pilot, it will be considered as a speedster and dropped from the sample.[7] Last, we plan to filter out respondents taking more than three-times the median time reported in the pilot.

A second attention check is related to comprehension of the choice card. The respondents are asked on the scope of the treatment, and are given the option between one correct answer and two wrong ones. Respondents that get this attention check wrong are allowed to continue answering the survey, but we will flag them in our final sample.

3.3 Pilot Data

Our study will be fielded in several stages. The first stage will contain answers from 150 respondents, the second stage will contain answers from 500, a third wave will contain 1500 completed surveys, and a fourth wave should contain the remaining completed questionnaires. We plan to use the first and second stages to adjust the level of our initial bids.

4 Analysis

We plan to analyze respondents' WTP using both standard and Latent Class models. In both cases, the dependent variable is the natural logarithm of WTP, which is interval-censored because it is elicited using double-bounded dichotomous-choice questions (Hanemann et al. 1991).

[7]For both the bad quality and speedsters responses, the survey company ensures a replacement of such respondents.

4.1 Statistical model

Let $\tilde{C}_i^*$ denote the elicited WTP, q_i denotes the baseline incidence, p_i denotes the baseline conditional mortality, h_i denote cancer's HRQL, d_i denote cancer's duration, CH_i denote cancer-year horizon, θ_{q_i} and θ_{p_i} denote reductions in incidence risk and conditional mortality risk, respectively. In addition, let $brd_i = q_i \times p_i$ denote the baseline risk of developing and dying from cancer, $brs_i = q_i \times (1 - p_i)$ the baseline risk of developing and recovering from cancer, $rrd_i = brs_i - (q_i - \theta_{q_i}) \times (p_i - \theta_{p_i})$ denote the reduced risk of developing and dying from cancer, $rrs_i = brr_i - (q_i - (\theta_{q_i}) \times (1 - (p_i - \theta_{p_i}))$ the reduced risk of developing and recovering from cancer. Finally, the remaining unobserved idiosyncratic variation is captured by ϵ_i.

All models will regress the natural logarithm of elicited WTP to the set of exogenous independent variables observed by respondents similar to the following model.

$$\log(C_i^*) = \alpha + \beta_{brq} brd_i + \beta_{brs} brs_i + \beta_h h_i + \beta_h d_i + \beta_{rrd} rrd_i + \beta_{rrs} rrs_i + \epsilon_i. \tag{1}$$

The independent variables will vary from model to model. Table 3 reports on the different specifications that we would like to run.

Table 3: Alternative specifications

	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**	**(9)**	**(10)**
Baseline										
brd_i	x	x				x	x			
brs_i	x	x				x	x			
q_i			x	x	x			x	x	x
p_i	x	x				x	x			
d_i	x	x	x	x	x	x	x	x	x	x
h_i	x	x	x	x	x	x	x	x	x	x
CH_i	x	x	x	x	x	x	x	x	x	x
Risk reduction										
θ_{q_i}	x	x				x	x			
θ_{p_i}	x	x				x	x			
rrs_i				x	x			x	x	
rrd_i				x	x			x	x	
$rrs_i + rrd_i$			x							x
Demographics (X_i)		x			x		x		x	
Type of model										
Log-Log	x	x	x	x	x					
Log-Lev						x	x	x	x	x

Models 1, 2, 3 regress the natural logarithm of elicited WTP to the set of exogenous independent variables observed by respondents on their choice cards. Model 6, 7 and 8, follow the settings from 1, 2 and 3, replacing the independent variables by their natural logarithms. Similarly, models 4, 5, and 9, 10 regress the natural logarithm of elicited WTP to the set of exogenous independent variables that are implied by the choice cards, but are not directly observed by respondents on their choice cards.
Finally, we plan to introduce a socio-demographic controls. The central demographics that we plan to include is household's income and household's number of children. Models 2 and 7 aim to report associations between respondents WTP and demographics. To attempt at controlling for the non-experimental (and self-reported) data on income, we plan to use as an exogenous shifter the average per capita income at the zip-code of residence.

4.2 Multiple outcome and multiple hypothesis testing

The coefficients of the models are to be estimated using maximum likelihood estimation (Alberini 1995) and the standard errors are to calculated using a Wald test (Train 2009). We will allow for correlation between errors within a respondent across choice occasions, but assume independence between respondents. Moreover, the standard errors from all combinations of parameters will be computed using the Delta Method.

4.3 Heterogeneous effects

We plan to use Latent class Analysis, as well as Moment Inequalities to allow for observed and unobserved heterogeneity.

5 Administrative information

5.1 Funding

Daniel Herrera-Araujo acknowledges financial support by the French National Research Agency (ANR) through the VHEALTH project, award number ANR-20-CE36-0010-01.

5.2 Institutional Review board

IRB Name: Comité d' Éthique de la Recherche de l'université Paris-Dauphine
IRB Approval Date: 10/07/2023
IRB Approval Number: 2023-04

5.3 Declaration of interest

The views expressed in this paper are those of the authors and do not represent official positions of the European Chemicals Agency. All authors report no conflit of interest.

5.4 Acknowledgements

We would like to thank Nicolas Koutsandreou, Adéle Lemoine, and Maria Koustopoulou for their excellent work as research assistants.

References

Alberini, A. (1995), 'Efficiency vs bias of willingness-to-pay estimates: bivariate and interval-data models', *Journal of environmental economics and management* **29**(2), 169–180.

Aldy, J. E. and Viscusi, W. K. (2007), 'Age differences in the value of statistical life: revealed preference evidence', *Review of Environmental Economics and Policy* .

Andrade, L. F., Ludwig, K., Goni, J. M. R., Oppe, M. and de Pouvourville, G. (2020), 'A french value set for the eq-5d-5l', *Pharmacoeconomics* **38**, 413–425.

Baker, R., Chilton, S., Donaldson, C., Jones-Lee, M., Lancsar, E., Mason, H., Metcalf, H., Pennington, M. and Wildman, J. (2011), 'Searchers vs surveyors in estimating the monetary value of a qaly: resolving a nasty dilemma for nice', *Health Economics, Policy and Law* **6**(4), 435–447.

Barsky, R. B., Juster, F. T., Kimball, M. S. and Shapiro, M. D. (1997), 'Preference parameters and behavioral heterogeneity: An experimental approach in the health and retirement study', *The Quarterly Journal of Economics* **112**(2), 537–579.

Carthy, T., Chilton, S., Covey, J., Hopkins, L., Jones-Lee, M., Loomes, G., Pidgeon, N. and Spencer, A. (1998), 'On the contingent valuation of safety and the safety of contingent valuation: Part 2-the cv/sg" chained" approach', *Journal of risk and uncertainty* **17**, 187–214.

Corso, P. S., Hammitt, J. K. and Graham, J. D. (2001), 'Valuing mortality-risk reduction: using visual aids to improve the validity of contingent valuation', *Journal of Risk and Uncertainty* **23**(2), 165–184.

Eckel, C. C. and Grossman, P. J. (2002), 'Sex differences and statistical stereotyping in attitudes toward financial risk', *Evolution and human behavior* **23**(4), 281–295.

Eckel, C. C. and Grossman, P. J. (2008), 'Forecasting risk attitudes: An experimental study using actual and forecast gamble choices', *Journal of Economic Behavior & Organization* **68**(1), 1–17.

Hammitt, J. K. (2013), 'Admissible utility functions for health, longevity, and wealth: integrating monetary and life-year measures', *Journal of risk and uncertainty* **47**, 311–325.

Hammitt, J. K. and Graham, J. D. (1999), 'Willingness to pay for health protection: inadequate sensitivity to probability?', *Journal of risk and uncertainty* **18**, 33–62.

Hammitt, J. K. and Haninger, K. (2010), 'Valuing fatal risks to children and adults: Effects of disease, latency, and risk aversion', *Journal of risk and Uncertainty* **40**, 57–83.

Hammitt, J. K. and Haninger, K. (2017), 'Valuing nonfatal health risk as a function of illness severity and duration: Benefit transfer using qalys', *Journal of Environmental Economics and Management* **82**, 17–38.

Hammitt, J. K. and Herrera-Araujo, D. (2018), 'Peeling back the onion: Using latent class analysis to uncover heterogeneous responses to stated preference surveys', *Journal of Environmental Economics and Management* **87**, 165–189.

Hanemann, M., Loomis, J. and Kanninen, B. (1991), 'Statistical efficiency of double-bounded dichotomous choice contingent valuation', *American journal of agricultural economics* **73**(4), 1255–1263.

Herrera-Araujo, D., Hammitt, J. K. and Rheinberger, C. M. (2020), 'Theoretical bounds on the value of improved health', *Journal of Health Economics* **72**, 102341.

Herrera-Araujo, D. and Rochaix, L. (2020), 'Does the value per statistical life vary with age or baseline health? evidence from a compensating wage study in france', *Journal of Environmental Economics and Management* **103**, 102338.

Krupnick, A. (2007), 'Mortality-risk valuation and age: stated preference evidence', *Review of Environmental Economics and Policy* .

Mason, H., Jones-Lee, M. and Donaldson, C. (2009), 'Modelling the monetary value of a qaly: a new approach based on uk data', *Health Economics* **18**(8), 933–950.

Pennington, M., Baker, R., Brouwer, W., Mason, H., Hansen, D. G., Robinson, A., Donaldson, C. and Team, E. (2015), 'Comparing wtp values of different types of qaly gain elicited from the general public', *Health economics* **24**(3), 280–293.

Pinto-Prades, J. L., Loomes, G. and Brey, R. (2009), 'Trying to estimate a monetary value for the qaly', *Journal of health economics* **28**(3), 553–562.

Pratt, J. W. and Zeckhauser, R. J. (1996), 'Willingness to pay and the distribution of risk and wealth', *Journal of Political Economy* **104**(4), 747–763.

Spiegelhalter, D. (2017), 'Risk and uncertainty communication', *Annual Review of Statistics and Its Application* **4**, 31–60.

Train, K. E. (2009), *Discrete choice methods with simulation*, Cambridge university press.

U.S. Department of Health and Human Services (HHS) (2016), 'Guidelines for regulatory impact assessment'.

Viscusi, W. K. and Aldy, J. E. (2003), 'The value of a statistical life: a critical review of market estimates throughout the world', *Journal of risk and uncertainty* **27**, 5–76.