Appendix E: Model, Estimation and Alternative Analyses

Table of Contents

1 Introduction ..................................................................................................................................... 3
2 Derivation of the estimating forms of the model ........................................................................... 3
  2.1 Development of the net income term ....................................................................................... 4
  2.2 Development of the health-state-related term .......................................................................... 5
  2.3 Development of the error term .................................................................................................. 6
  2.4 The difference in discounted expected utilities that drives choices ........................................ 7
  2.5 Solving for $WTP$ .................................................................................................................... 8
    2.5.1 The linear form .................................................................................................................. 8
    2.5.2 Normalizations .................................................................................................................. 9
    2.5.3 The Box-Cox transformation for net income .................................................................... 10
    2.5.4 The shifted logarithmic transformation for health states .............................................. 10
  2.6 Simulated distributions for $WTP$ .......................................................................................... 11
3 Estimation ..................................................................................................................................... 12
  3.1 Panel data: Fixed Effects? ....................................................................................................... 12
  3.2 Panel data: Random Parameters? ........................................................................................... 13
4 Fixed effects versus no fixed effects ............................................................................................. 13
  4.1 Biostatistical Perspective ........................................................................................................ 13
  4.2 Econometric Perspective ......................................................................................................... 14
  4.3 Hausman test for fixed effects ................................................................................................. 17
5 Random-parameters logit models .................................................................................................. 18
  5.1 Results: Random parameters specifications .......................................................................... 19
6 Alternate Specifications .................................................................................................................. 19
  6.1 Preliminary models .................................................................................................................. 19
  6.2 Appropriate transformation for health state durations ............................................................ 20
  6.3 Different assumptions about the individual discount rate ...................................................... 20
  6.4 Including an alternative-specific dummy for “either program” ............................................. 21
  6.5 If respondents anticipate having only half as much income when sick .................................. 22
  6.6 If respondents perceive other costs in addition to those quoted ............................................ 22
7 Tables .......................................................................................................................................... 27
  7.1 Table E1 – Preliminary simple models .................................................................................... 27
  7.2 Table E1 – Different assumptions about the discount rate .................................................... 28
1 Introduction
In this appendix, we provide detailed descriptions of how we arrive at the estimating specifications used in our model. We also provide an outline of how Stata’s fixed effects conditional logit model is specified and estimated.

2 Derivation of the estimating forms of the model
We suppress the $i$ subscripts for individuals and write indirect utility levels in each future period as follows.

\begin{align*}
V_{i}^{AH} &= f(\text{net}Y_{i}) + \epsilon_{i}^{AH} \\
V_{i}^{AS} &= f(\text{net}Y_{i}) + \alpha_{1}(\text{illness}_{i}) + \alpha_{2}(\text{recovered}_{i}) + \alpha_{3}(\text{lost life-year}_{i}) + \epsilon_{i}^{AS} \\
V_{i}^{NH} &= f(\text{net}Y_{i}) + \epsilon_{i}^{NH} \\
V_{i}^{NS} &= f(\text{net}Y_{i}) + \alpha_{1}(\text{illness}_{i}) + \alpha_{2}(\text{recovered}_{i}) + \alpha_{3}(\text{lost life-year}_{i}) + \epsilon_{i}^{NS}
\end{align*}

In future period $t$, the difference in expected utility with program A and with no program (N):

\begin{equation}
\left[ (1 - \Pi^{AS})V_{t}^{AH} + (\Pi^{AS})V_{t}^{AS} \right] - \left[ (1 - \Pi^{NS})V_{t}^{NH} + (\Pi^{NS})V_{t}^{NS} \right]
\end{equation}

To explain a decision taken today, based on the stream of future differences in expected indirect utilities across the two alternatives, these future quantities must be discounted back to the present.

The fact that net income and health status are assumed to be approximately level within each of the four different health states permits us to reverse the order of discounting and the taking of expectations. We can work in terms of the present discounted time in each health state, and simply multiply this by the utility of net income in that interval and by the (dis)utility of health status in that interval. We assume simple exponential discounting where the discount factor is $\delta^{t} = (1 + r)^{-t}$. Each summation in the following terms runs from the present to the end of the individual’s nominal lifespan.

\begin{align*}
pdve &= \sum \delta^{t}1(\text{pre-illness}_{i}) \\
pdvi &= \sum \delta^{t}1(\text{illness}_{i}) \\
pdvr &= \sum \delta^{t}1(\text{recovered}_{i}) \\
pdvl &= \sum \delta^{t}1(\text{lost life-year}_{i})
\end{align*}

For convenience, we define to other types of present discounted time intervals,

$pdvp = pdve + pdvr$, just the time where the individual is neither sick nor dead, and

$pdvc = pdve + pdvi + pdvr + pdvl$, the entire remainder of the individual’s nominal lifespan.

We will now develop, separately, the “present discounted expected” form of the three parts of the indirect utility function: the net income terms, the health status terms, and the error...
term. Fortunately, we find no evidence that the marginal utility of net income depends on these probabilistic future health states (or vice versa) in any of the models explored in the main paper. In other work, we find some evidence of the dependence of the marginal utility of net income on current health, but this is the numeraire health state in the main paper.

2.1 Development of the net income term

The net income level, $net_i$, will differ according to the type of health state, whether the program is currently being paid for, and whether the individual gets sick or stays healthy:

<table>
<thead>
<tr>
<th>Indirect utility, Probability</th>
<th>Pre-illness/latency (“c”)</th>
<th>Illness/injury time (“i”)</th>
<th>Recovered/remission (“r”)</th>
<th>Lost life-years (“l”)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$V_i^{AH}, (1 - \Pi_i^{AS})$</td>
<td>$Y_c$</td>
<td>$Y_c$</td>
<td>$Y_c$</td>
<td>$Y_c$</td>
</tr>
<tr>
<td>$V_i^{AS}, \Pi_i^{AS}$</td>
<td>$Y_c$</td>
<td>$Y$</td>
<td>$Y_c$</td>
<td>$0$</td>
</tr>
<tr>
<td>$V_i^{NH}, (1 - \Pi_i^{NS})$</td>
<td>$Y$</td>
<td>$Y$</td>
<td>$Y$</td>
<td>$Y$</td>
</tr>
<tr>
<td>$V_i^{NS}, \Pi_i^{NS}$</td>
<td>$Y$</td>
<td>$Y$</td>
<td>$Y$</td>
<td>$0$</td>
</tr>
</tbody>
</table>

Discounted time in health state: $pdve$, $pdvi$, $pdvr$, $pdvl$

We can make use of our notation for discounted future time intervals, plus the pattern of net income amounts under the four different outcomes as displayed in the table above, to write the discounted future expected utility from net income as follows (noting that $pdve + pdvi + pdvr = pdvc - pdvl$):

$$
\left[ (1 - \Pi_i^{AS} ) f (Y - c) pdvc \\
+ (\Pi_i^{AS} ) \left[ f (Y - c) pdvp + f (Y) pdvi \right] \\
- \left[ (1 - \Pi_i^{NS} ) f (Y) pdvc \\
+ (\Pi_i^{NS} ) f (Y)(pdvc - pdvl) \right] \right]
$$

(4)

Distribute the probabilities, collect terms, and rearrange in the following steps:

$$
f (Y - c) pdvc - \Pi_i^{AS} f (Y - c) pdvc + \Pi_i^{AS} f (Y - c) pdvp + \Pi_i^{AS} f (Y) pdvi \\
- f (Y) pdvc + [\Pi_i^{NS} f (Y) pdvc - \Pi_i^{NS} f (Y) pdvc] + \Pi_i^{NS} f (Y) pdvl
$$

(5)

Gathering the terms in $f (Y - c)$ and $f (Y)$ and simplifying allows equation (4) to be written in a couple of different ways, for example as:
\[
f(Y - c) \left[ (1 - \Pi^{AS}) pdvc + \Pi^{AS} pdvp \right] \\
- f(Y) \left[ (1 - \Pi^{NS}) pdvc + \Pi^{NS} pdvp - \left( \Pi^{AS} - \Pi^{NS} \right) pdvi \right]
\]  
(6)

or as:

\[
f(Y - c) \left[ \Pi^{AH} pdvc + \Pi^{AS} pdvp \right] \\
- f(Y) \left[ \Pi^{NH} pdvc + \Pi^{NS} pdvp - \Delta \Pi^{AS} pdvi \right]
\]  
(7)

Either way, we denote the terms in square brackets using further abbreviations:

\[
cterms = (1 - \Pi^{AS}) pdvc + \Pi^{AS} pdvp \\
= \Pi^{AH} pdvc + \Pi^{AS} pdvp
\]  
\[
yterms = pdvc - \Pi^{AS} pdvi - \Pi^{NS} pdvl \\
= \Pi^{NH} pdvc + \Pi^{NS} pdvp - \Delta \Pi^{AS} pdvi
\]  
(8)

2.2 Development of the health-state-related term

We assume that our subjects view future health states when “healthy” or “sick” as being the same, regardless of whether Program A or No Program is selected. All that is affected by Program A is the risk of suffering this illness profile. Unlike the net income profiles, therefore, the “net health” profile over time depends only on whether the individual gets sick.

<table>
<thead>
<tr>
<th>Utility from one period in each health state, as a function of program choice and health outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td>Indirect utility, Probability</td>
</tr>
<tr>
<td>(V_t^{AH}, (1 - \Pi^{AS}))</td>
</tr>
<tr>
<td>(V_t^{AS}, \Pi^{AS})</td>
</tr>
<tr>
<td>(V_t^{NH}, (1 - \Pi^{NS}))</td>
</tr>
<tr>
<td>(V_t^{NS}, \Pi^{NS})</td>
</tr>
<tr>
<td>Discounted time in health state:</td>
</tr>
</tbody>
</table>

Written in its extensive form the difference in discounted expected health states between Program A and no program is given by:
\[
\begin{align*}
(1 - \Pi^{AS}) & \left[ \alpha_0 pdve + \alpha_0 pdvi + \alpha_0 pdvr + \alpha_0 pdvl \right] \\
+ (\Pi^{AS}) & \left[ \alpha_0 pdve + \alpha_1 pdvi + \alpha_2 pdvr + \alpha_3 pdvl \right] \\
- (1 - \Pi^{NS}) & \left[ \alpha_0 pdve + \alpha_0 pdvi + \alpha_0 pdvr + \alpha_0 pdvl \right] \\
+ (\Pi^{NS}) & \left[ \alpha_0 pdve + \alpha_1 pdvi + \alpha_2 pdvr + \alpha_1 pdvl \right]
\end{align*}
\]

Distributing the probability terms and simplifying yields:

\[
(\Pi^{AS} - \Pi^{NS}) \begin{bmatrix}
(\alpha_0 - \alpha_0) pdve \\
+ (\alpha_1 - \alpha_0) pdvi \\
+ (\alpha_2 - \alpha_0) pdvr \\
+ (\alpha_3 - \alpha_0) pdvl
\end{bmatrix}
\]

If we normalize future health-related utility on the individual’s status quo health state, equivalent to setting $\alpha_0 = 0$, and express the change in the risk of the illness profile due to Program A as $\Delta \Pi^{AS} = \Pi^{AS} - \Pi^{NS}$, we can write this term more simply as:

\[
(\alpha_1 pdvi + \alpha_2 pdvr + \alpha_3 pdvl) \Delta \Pi^{AS}
\]

Here, the estimated $\alpha$ parameters are the (dis)utilities from one unit of time in each adverse health state, relative to the individual’s current pre-illness health status. This normalization is particularly convenient. However, it imposes some strong assumptions which we explore in other work, where we allow these marginal disutilities of adverse future health states to depend upon current morbidities and comorbidities, and upon subjective risks for the health problem in question and other major types of health risks. The marginal disutilities estimated in our basic models must be interpreted as averages, across the current population distribution of health states and health outlooks, for the U.S. population 25 years and older.

2.3 Development of the error term

For completeness, the assumed independent and identically error terms in each of the four variants of indirect utility in each future period are combined in a similar fashion:

\[
\left[ (1 - \Pi^{AS}) \varepsilon_i^{AH} + (\Pi^{AS}) \varepsilon_i^{AS} \right] - \left[ (1 - \Pi^{NS}) \varepsilon_i^{NH} + (\Pi^{NS}) \varepsilon_i^{NS} \right]
\]

When discounted back to the present, we assume the resulting differences in expected error terms (across the healthy and sick outcomes) are cooperative in being distributed in a manner consistent with the assumptions necessary for the use of McFadden’s conditional logit choice model.
2.4 The difference in discounted expected utilities that drives choices

We can now assemble the discounted net income terms, the discounted health state terms, and the discounted error terms to yield the difference in discounted expected utilities that is assumed to drive the individual’s choice between Program A and “No program.”

\[
\Delta PDV (E[V]) = \{f(Y-c)_{cterm} - f(Y)_{vterm}\} \\
+ [\alpha_1 pdvi + \alpha_2 pdvr + \alpha_3 pdvl] \Delta \Pi^{AS} + \varepsilon
\] (13)

This is the basis for the estimating equation used in the paper. Generalization to the case of three alternatives simply means we introduce a second “difference” equation analogous to equation (13), but for risk reduction Program B, relative to “No Program.” Program costs and the size of the risk reduction, as well as the relevant illness profile, will differ between the two programs. For the “Neither program” alternative, of course, the “difference relative to Neither Program” is zero for all variables. There is no difference in net income because program costs are not incurred, and the term involving the health profile is zero because there is no reduction in the risk of experiencing that profile (i.e. \(\Delta \Pi^{BS} = 0\)). The health risk is still present, but since neither program is selected, no reduction in risk is achieved.

All that remains is to choose a specific functional form for \(f(\cdot)\) and to decide whether preferences are homogeneous or whether the data suggest that they should be specified as heterogeneous (i.e. a function of observable individual attributes). In the main paper, Cameron and DeShazo (2009), we eventually depart from this model based directly on future per-period health state utilities. Instead, we allow individuals’ decisions to be based directly on “present discounted time in future adverse health states” as the proximal determinants of choice. We consider nonlinear forms in \(pdvi\), \(pdvr\), and \(pdvl\), and find that a flexible translog-type functional form seems to provide the best fit to the choice data.

The systematic portion of equation (13), provided it can be written as a linear-in-parameters function of variables constructed from our data, can be interpreted as the \(x \beta\) term in the standard conditional logit (and fixed effects conditional logit) models that we use to estimate the parameters of our models. The data also suggest that the function \(f(\cdot)\) should be nonlinear. We have explored quadratic forms, square root forms, and Box-Cox-type forms with a transformation parameter of 0.42, determined via a line-search. The quadratic form is the most general, but involve one more parameter and permits marginal utility to go negative at extreme values of net income in some models with heterogeneous marginal utilities. The square root form is very close to the Box-Cox transformation with a parameter of 0.5, but reviewers of our early results have suggested that the 0.42 parameter may be preferable. In the main paper, we treat this parameter as a known constant, rather than estimating it using a fixed effects conditional logit model with a nonlinear-in-parameters “index” \((x \beta\) term) since such a model is not readily available. Treating this value of the parameter as fixed is certainly no worse than using a linear or logarithmic specification and implicitly assuming a Box-Cox transformation parameter that is fixed at one or zero. In other work, we are developing models which permit nonlinearities in the logit index, in particular to accommodate estimated values of the discounting parameter, which we also treat as fixed in the main paper, although we explore the consequences of alternative assumptions.
2.5 Solving for WTP

Once the indirect utility parameters have been estimated using respondents’ choices, the next step is to solve for the value of \( c \) that makes the individual just indifferent between paying for the program and getting the benefits, or not paying for the program and doing without the benefits. Recall that \( c \) is the annual payment for the risk-reduction program, assumed to be paid only if the individual is not currently afflicted by the illness in question or prematurely dead from the same illness.

In general, equation (13) can be solved for this maximum annual willingness to pay (while not sick or dead):

\[
0 = f \left( Y - c^* \right)_{\text{term}} - f \left( Y \right)_{\text{term}} + \left[ \alpha_1 p_{dvi} + \alpha_2 p_{dvr} + \alpha_3 p_{dvl} \right] \Delta \Pi^{4S} + \epsilon
\]

\[
- f \left( Y - c^* \right)_{\text{term}} = f \left( Y \right)_{\text{term}} + \left[ \alpha_1 p_{dvi} + \alpha_2 p_{dvr} + \alpha_3 p_{dvl} \right] \Delta \Pi^{4S} + \epsilon
\]

\[
f \left( Y - c^* \right)_{\text{term}} = \frac{-1}{f \left( Y \right)_{\text{term}}} f \left( Y \right)_{\text{term}} + \left[ \alpha_1 p_{dvi} + \alpha_2 p_{dvr} + \alpha_3 p_{dvl} \right] \Delta \Pi^{4S} + \epsilon
\]

\[
Y - c^* = f^{-1} \left( f \left( Y \right)_{\text{term}} + \left[ \alpha_1 p_{dvi} + \alpha_2 p_{dvr} + \alpha_3 p_{dvl} \right] \Delta \Pi^{4S} + \epsilon \right)
\]

\[
c^* = Y - f^{-1} \left( f \left( Y \right)_{\text{term}} + \left[ \alpha_1 p_{dvi} + \alpha_2 p_{dvr} + \alpha_3 p_{dvl} \right] \Delta \Pi^{4S} + \epsilon \right)
\]

2.5.1 The linear form

It can be helpful to consider the WTP formula if the income function is \( f \left( Y \right) = \beta Y \), since this simpler form aids in developing a clear intuition about the determinants of WTP. For the \( i \)th individual and the illness profile for that individual that would be addressed by Program A, which would reduce the risk of getting sick with that illness by \( \Delta \Pi_{i}^{4S} \), the maximum annual WTP during non-sick and non-dead years will be:

\[
c_{i}^{4*} = Y_{i} - \frac{1}{\beta} \left( \beta Y_{i} \text{term}_{i}^{A} - \alpha_1 \left( \Delta \Pi_{i}^{4S} \text{pdvi}_{i}^{A} \right) - \alpha_2 \left( \Delta \Pi_{i}^{4S} \text{pdrv}_{i}^{A} \right) - \alpha_3 \left( \Delta \Pi_{i}^{4S} \text{pdvl}_{i}^{A} \right) - \epsilon_{i}^{4} \right)_{\text{term}_{i}^{A}}
\]

\[
\text{PDV} \left( E \left[ c_{i}^{4*} \right] \right) = \left( \left( 1 - \Pi^{4S} \right) p_{dvc} + \Pi^{4S} p_{dvp} \right) c_{i}^{4*}
\]

\[
= \left( \text{term}_{i}^{A} \right) c_{i}^{4*}
\]

The first term in the braces is the chance of staying healthy under the program, time the present discounted number of years left in the individual’s nominal lifespan. The second term is the...
chance of suffering the illness, time the discounted number of years when neither sick nor dead. The fact that the term in braces is the same as the cterm expression used in the formula for $c^*_i$ is very convenient, since this aids us in simplifying the formula.

$$\text{PDV}\left( E\left[ \hat{c}^*_i \right] \right) = cterm^*_i \left[ Y_i - \frac{1}{\beta} \left( \beta Y_i \text{term}^*_i - \left[ \alpha_1 pdvi^*_i + \alpha_2 pdvr^*_i + \alpha_3 pdvl^*_i \right] \Delta \Pi^*_i - \varepsilon^*_i \right) \right]$$

$$= Y_i \left( cterm^*_i - \text{term}^*_i \right) + \frac{\alpha_1}{\beta} \Delta \Pi^*_i pdvi^*_i + \frac{\alpha_2}{\beta} \Delta \Pi^*_i pdvr^*_i + \frac{\alpha_3}{\beta} \Delta \Pi^*_i pdvl^*_i + \frac{\varepsilon^*_i}{\beta}$$

(17)

$$= -Y_i \left( \Delta \Pi^*_i pdvl^*_i \right) + \frac{\alpha_1}{\beta} \Delta \Pi^*_i pdvi^*_i + \frac{\alpha_2}{\beta} \Delta \Pi^*_i pdvr^*_i + \frac{\alpha_3}{\beta} \Delta \Pi^*_i pdvl^*_i + \frac{\varepsilon^*_i}{\beta}$$

(18)

### 2.5.2 Normalizations

#### 2.5.2.1 Normalization on an aggregate 1.00 risk change (a “statistical life”)

The next step is to normalize the WTP amount on some arbitrary-sized risk reduction. In the literature on VSLs, the convention is to normalize WTP on a 1.00 risk change, which involves scaling up the WTP estimate proportionately. As a practical matter, this is done by dividing the equation in (17) by the size of the risk change conferred by Program A:

$$\frac{\text{PDV}\left( E\left[ \hat{c}^*_i \right] \right)}{\Delta \Pi^*_i} = -Y_i \left( pdvl^*_i \right) + \frac{\alpha_1}{\beta} pdvi^*_i + \frac{\alpha_2}{\beta} pdvr^*_i + \frac{\alpha_3}{\beta} pdvl^*_i + \frac{\varepsilon^*_i}{\beta \Delta \Pi^*_i}$$

(18)

This risk change, however, is a negative number. If we wish to think in terms of a positive-sized risk reduction, of size $|\Delta \Pi^*_i|$, we could divide through, instead, by the absolute value of the risk reduction $|\Delta \Pi^*_i|$, yielding the alternative formula where all terms on the right-hand side will have the opposite sign:

$$\frac{\text{PDV}\left( E\left[ \hat{c}^*_i \right] \right)}{|\Delta \Pi^*_i|} = Y_i pdvl^*_i - \frac{\alpha_1}{\beta} pdvi^*_i - \frac{\alpha_2}{\beta} pdvr^*_i - \frac{\alpha_3}{\beta} pdvl^*_i - \frac{\varepsilon^*_i}{\beta |\Delta \Pi^*_i|}$$

(19)

Bear in mind that the parameters $\alpha_1$, $\alpha_2$, and $\alpha_3$ are expected to be negative, since a greater number of discounted years in each of these adverse health states is expected to decrease the individual’s utility level. Thus for a linear-in-income version of the model in equation (13), individuals can be expected to be willing to pay a greater amount to avoid a particular illness profile, the greater their income and the greater the number of discounted years in adverse health states.
We seek only to describe the *expected* WTP for a risk reduction for a given illness profile, by a particular type of individual (rather than to predict any individual value). Thus we set the error term $\varepsilon_i$ to its expected value of zero and ignore the scale of its variance, given in the denominator of the last term in equation (19).

### **2.5.2.2 Normalization on a “microrisk reduction”**

The formula in equation (19) will produce something analogous to the value of a statistical life ($VSL$), which can be expected to be on the order of millions of dollars for illness profiles comparable to sudden death in the current period. In the main paper, however, we argue for normalization on a micro-risk reduction. This is achieved by dividing through not by the absolute size of the risk change, but by this risk change normalized on 0.000001.

\[
WTP_{\text{microrisk}} = \frac{PDV\left(E\left[\hat{\Delta}^A_i\right]\right)}{|\Delta \Pi_i^{AS}|/0.000001} = \frac{PDV\left(E\left[\hat{\Delta}^A_i\right]\right)}{|\Delta \Pi_i^{AS}|}(0.000001) \tag{20}
\]

### **2.5.3 The Box-Cox transformation for net income**

For a Box-Cox transformation of net income, the relevant formula for the maximum annual willingness to pay while neither sick nor dead is:

\[
e^* = Y - \left(1 + \lambda \left\{ -1 \frac{\lambda}{\hat{\lambda}} Y^{\lambda - 1} \right\} y_{\text{term}} + \left[\alpha_1 p_{dvi} + \alpha_2 p_{dvr} + \alpha_3 p_{dvl}\right] \Delta \Pi^{AS} + \varepsilon \right)^{1/\hat{\lambda}} \tag{21}
\]

This Box-Cox transformation of net income is the function used in the main paper, for a value of $\hat{\lambda} = 0.42$. (This value was determined via a line search. To estimate $\hat{\lambda}$ simultaneously with the other indirect utility parameters would require original programming in generalized nonlinear optimization software. There is a premium on forms that lend themselves to a linear-in-parameters “index” for the estimating specification so that packaged software can be used.) By following steps analogous to those used in the linear-in-income case in the last section, it is straightforward (if a little more tedious) to arrive at formulas for the WTP for a microrisk reduction in this case.

### **2.5.4 The shifted logarithmic transformation for health states**

In the main paper, we determined early in our analysis that the portion of equation (21) in square brackets that characterizes the illness profile (i.e. the discounted years in each adverse health state) is too restrictive. The data support a nonlinear specification in discounted health-state years. We considered more general forms, exploring a common Box-Cox transformation shared by each of these three terms (after adding one to all values). The effect on the maximized log-likelihood function is shown in Figure E2. Between a linear form and the shifted logarithmic transformation, the latter is more appropriate for these data. Only minor changes in the log-likelihood occur as the transformation parameter is reduced below the implicit zero value that is associated with the logarithmic transformation.
In addition to switching to the logarithmic transformation, we explored a full set of second-order terms. The higher-order terms in lost life-years were robustly significant, as was an interaction term between sick-time and lost life-years. Thus we retain these terms.

Finally, since there is considerable heterogeneity by age in the types of illness profiles our respondents were invited to consider, it is important to control for age in these specifications. We thus replace the three terms in square brackets in equation (21) with:

\[
\left[ \alpha_{10} \log \left( pdvi_{it} + 1 \right) + \alpha_{20} \log \left( pdvl_{it} + 1 \right) + \alpha_{30} \log \left( pdvl_{it} + 1 \right) + \alpha_{40} \log \left( pdvl_{it} + 1 \right) \right] 
+ \alpha_{41} \text{age}_{it} \left[ \log \left( pdvl_{it} + 1 \right) + \log \left( pdvl_{it} + 1 \right) \right] 
+ \alpha_{42} \text{age}_{it} \left[ \log \left( pdvl_{it} + 1 \right) + \log \left( pdvl_{it} + 1 \right) \right]
\]

There is an additional interaction term involving the first sick-years term, to help correct for selection bias in the parameter estimates. This selectivity-related term is described in Appendix D. See section 6.2 of this appendix for a justification of the choice of a shifted logarithmic transformation for each of the discounted health state terms.

2.6 Simulated distributions for WTP

After estimating the parameters of equation (13) using maximum likelihood methods for discrete choice (discussed in the section below on Estimation), the point estimates for each of the parameters could be substituted into the formula for WTP for a microrisk reduction to yield a fitted WTP for each program offered to each individual. However, the range of programs used in our stated preference survey instrument does not represent the distribution of health risks in the real world. These stylized and hypothetical health risks and the hypothetical programs proposed for reducing these risks are essential to the task of learning about consumer preferences, but that is the limit of our interest in them. However, once we have estimates for the preference parameters, we are interested in applying them to “real” illness profiles.

The main paper outlines how the model could be employed with the range of illness profiles and the types of affected individuals in a real policy context. For the paper, however, we pick just a handful of specific illustrative cases, each with a single specified illness profile that affects a single specified person. Point estimates for the WTP for a microrisk reduction for an illness profile could be obtained by substituting the point estimates of the indirect utility parameters into the appropriate formula for WTP. However, this would ignore the fact that the point estimates are not the true parameters, just estimates of those parameters (which are random variables). A better picture of the predictive capability of the estimated model can be obtained by simulating a distribution for the WTP amount.

We simulate WTP amounts by taking 1000 random draws from the asymptotic joint normal distribution of the maximum likelihood parameter estimates. For each draw, we calculate the corresponding value of the WTP for a microrisk reduction. Across all 1000 draws,

1 The “drawnorm” utility in Stata is very useful for this type of exercise.
we build a “sampling distribution” for the \( WTP \) amount that reflects the variability in the estimated indirect utility parameters. In the tables in the paper, we report the empirical median, and the empirical 5\(^{th}\) and 95\(^{th}\) percentiles of this distribution of 1000 values. This information provides a sense of the central tendency and the dispersion of the quantity of interest: \( WTP \) for a risk reduction.

As always, when simulating such distributions, it is readily apparent that there is a tradeoff between bias and efficiency. While it may be tempting to include less statistically significant explanatory variables in the choice model, the presence of insignificant coefficients can inflate the variance-covariance matrix for the parameters and this can result in very wide 90 percent intervals for the simulated distribution of \( WTP \) for a microrisk reduction. We typically worry that omission of insignificant variables may incur some degree of bias in the estimates of the remaining indirect utility parameters. However, given the extent of the randomization of the illness profiles and program costs in this study, there is little concern about omitted variables bias. Parsimonious specifications are entirely appropriate.

3 Estimation

Now we provide some background concerning the estimators used to produce the vector of maximum likelihood parameter point estimates and the parameter asymptotic variance-covariance matrix reported in the paper. This discussion assumes that the reader is familiar with conventional textbook treatments of models for unordered multiple discrete choice, for example, as covered in Greene (2008, section 23.11).

The choice sets faced by each respondent on each choice occasion in our study consist of three alternatives: Program A, Program B, or Neither Program. The “dependent variable” in this context is actually a set of three indicator variables, switched “on” if the corresponding alternative is chosen, and “off” if it is not. The explanatory variables all differ across alternatives: net income will depend upon which alternative is chosen because each program has a different cost. The chance of suffering each featured illness profile, interacted with the nature of that illness profile, will also differ across all three alternatives. Thus each explanatory variable differs both across individuals and across alternatives within each choice set the individual faces, so the “conditional” logit model is appropriate, as in Greene (2008, section 23.11.2, p. 846-847). Our constructed explanatory variables, as featured in our Table 3, are examples of Greene’s \( x_{ij} \) variables.

3.1 Panel data: Fixed Effects?

The first notable thing about the structure of our data on respondents’ three-way multiple discrete choices is that these are effectively “panel” data. Each respondent, typically, provides us with five different choices. With panel data, there is always a question whether a set of slope coefficients, estimated using simply the pooled data without recognition of its panel nature, might be affected by heterogeneity bias. (Heterogeneity bias is a form of omitted variables bias, where the explicit explanatory variables are correlated with unobserved forms of heterogeneity across individuals, so that the estimated slope coefficients are biased). Fortunately, the randomized design of all of our choice sets, conditional only on the age and gender of the respondent and the plausibility of some types of outcomes, means that the \( x_{ij} \) variables in our models are unlikely to be correlated with any omitted variables, especially since we control for the respondent current age in our models.
Nevertheless, the fact that we have repeated choices for each person in our sample immediately led us (and almost every other reviewer of our work) to a concern that appropriate panel-oriented econometric methods should be used with these data. The parameters of our model are thus estimated using the fixed effects conditional logit choice model as implemented in the Stata 10 econometric software package. The model is described in considerable detail in the Stata 10 Reference Manual under the heading “clogit – Conditional (fixed effects) logistic regression” (p. 285-287).

3.2 Panel data: Random Parameters?

A further possibility is that our choice models should be estimated using random-parameters logit models. These model permits each utility parameter to be individual-specific and the same across all five choices made by the same individual. However, it assumes that these individual-specific parameters are a random draw from a joint distribution of utility parameters in the population. The goal is to estimate both a central tendency and a dispersion for each utility parameter, to allow explicitly for unobserved forms of heterogeneity in preferences.

In Section 2, we describe first the biostatistical version of the fixed-effects logit model, as summarized in the documentation for Stata’s algorithms. In Section 3, we then provide an alternative description of these models, from an econometric perspective, as explained in Greene (2008). In Section 4, we describe the results of using random parameters (mixed) logit models with our data, along with our rationale for preferring to estimate systematically varying parameters, rather than randomly varying parameters.

4 Fixed effects versus no fixed effects

Breslow and Day (1980, 247-279), Collett (2003, 251-267), and Hosmer and Lemeshow (2000, 223-259) provide the biostatistics version of “conditional logistic regression.” Hamerle and Ronning (1995) also describe the fixed-effects logit, but Chamberlain (1980) is the standard econometrics reference for this model. We provide both the biostatistical and the econometric perspectives on this model in the two sections to follow:

4.1 Biostatistical Perspective

We use the pre-programmed algorithms in the Stata software package to estimate our fixed effects logit models. Stata’s description of the estimator is couched in terms of the biostatistical approach to these models. For those who are most familiar with that approach, we adapt the description in the Stata manual, tailoring it to the application of the model in this paper, and using conformable notation, let \( i = 1, \ldots, n \) denote respondents and let \( k = 1, \ldots, 5 \) denote the five choice scenarios presented to each respondent. We will start with an exposition which assumes just the choice between “Program A” and “No program” (N). Let \( y_{ik} \) be the dependent variable taking on values 1 if the program is chosen and 0 if no program is chosen. Let \( y_i = (y_{i1}, \ldots, y_{i5}) \) be the outcomes for the \( i^\text{th} \) respondent. Let \( x_{ik} \) be a row vector of covariates (i.e. the explanatory variables listed as regressors in Table 3 in the paper). Let

\[
h_{it} = \sum_{k=1}^{5} y_{ik}
\]  

(23)
be the observed number of ones for the dependent variable for the \( i^{th} \) respondent. In the biostatistical version of the model, practitioners would say that there are \( h_{li} \) “cases” matched to \( h_{2i} = 5 - h_{li} \) “controls” for the \( i^{th} \) respondent.

In the analysis, we consider the probability of a possible value of \( y_i \), the vector of outcomes, conditional on \( \sum_{k=1}^{5} y_{ik} = h_{li} \) (Hamerle and Ronning, 1995, equation 8.33; Hosmer and Lemeshow, 2000, equation 7.4),

\[
\Pr(y_i | \sum_{k=1}^{5} y_{ik} = h_{li}) = \frac{\exp\left(\sum_{k=1}^{5} y_{ik} x_{ik} \beta\right)}{\sum_{d_k=0}^{1} \exp\left(\sum_{k=1}^{5} d_k x_{ik} \beta\right)}
\]

where \( d_k \) is equal to 0 or 1 with \( \sum_{k=1}^{5} d_k = h_{li} \) and \( S_i \) is the set of all possible combinations of \( h_{li} \) ones and \( h_{2i} \) zeros. There are \( \binom{5}{h_{li}} \) such combinations, but we fortunately do not need to count all these combinations to compute the denominator in equation (24), since it can be computed recursively. Denote the needed denominator as:

\[
g_i(5, h_{li}) = \sum_{d_k=0}^{1} \exp\left(\sum_{k=1}^{5} d_k x_{ik} \beta\right)
\]

Consider, computationally, how \( g_i \) changes as we go from a total of 1 choice set per person to 2 choice sets, and so on. Doing this, we derive the recursive formula:

\[
g_i(5, h) = g_i(4, h) + g_i(4, h-1) \exp(x_i \beta)
\]

where we define \( g_i(5, h) = 0 \) if \( 5 < h \) and \( g_i(5, 0) = 1 \).

The conditional log-likelihood for this problem is:

\[
\ln L = \sum_{i=1}^{n} \left\{ \sum_{k=1}^{5} y_{ik} x_{ik} \beta - \log g_i(5, h_{li}) \right\}
\]

where the derivatives of the conditional log-likelihood can also be computed recursively by taking derivatives of the recursive formula for \( g_i \).

The documentation for Stata 10 indicates that computation time is roughly proportional to \( p^2 \sum_{i=1}^{n} \min(h_{li}, h_{2i}) \), where \( p \) is the number of independent variables in the model. If \( \min(h_{li}, h_{2i}) \) is small, computation time is not an issue.

### 4.2 Econometric Perspective

Based on Greene (6e, 2008), Ch. 23.5.2, and the references cited therein), we can adapt the discussion of the choice probabilities employed in Chamberlain’s conditional likelihood function.
to a simple case which conveys the intuition of the fixed effects logit approach. Suppose the systematic portion of the indirect utility differences associated with each individual include a component that is constant for any one individual but differs across individuals: $\alpha_i + x_{ik}\beta$. In the context of the models explored in our main paper, the $x_{ik}$ vector consists of the thirteen explanatory variables constructed from our raw data and featured in Model 5 in Table 3, where the model involves five choice scenarios per person, each concerning three alternatives. To keep the algebra simple, consider the binary choice case, rather than the three-way choice case, with the recognition that it can be generalized to the three-alternative case considered in the body of our paper.

The unconditional likelihood function, when there are $K$ choices involving just a pair of alternatives for each individual, will take the following form, where the regressors, $x_{ik}$, are implicitly the differences between the attributes of the two alternatives between the “1” and the “0” outcome (often the status quo outcome for which attribute levels are normalized to zero):

$$L = \prod_{i=1}^{n} \prod_{k=1}^{K} \left( \frac{\exp(\alpha_i + x_{ik}\beta)}{1 + \exp(\alpha_i + x_{ik}\beta)} \right)^{y_{ik}} \left( \frac{1}{1 + \exp(\alpha_i + x_{ik}\beta)} \right)^{1-y_{ik}} \quad (28)$$

For the five different three-way choices made by respondents in our study, the corresponding unconditional likelihood function would involve three distinct indicators, $y_{Aik}$, $y_{Bik}$, and $y_{Nik}$ that take the value 1 if the corresponding alternative among A, B, and N is chosen, and the value 0 otherwise. The regressors are the attributes associated with each alternative, normalized on their levels for the “Neither Program” alternative to permit estimation of a unique parameter vector. This would be an ordinary pooled-data conditional logit model except for the individual-specific constant which shifts the systematic utility associated with every non-numeraire alternative.

$$L = \prod_{i=1}^{n} \prod_{k=1}^{5} \left[ \frac{\exp(\alpha_i + (x_{Aik} - x_{Nik})\beta)}{\exp(\alpha_i + (x_{Aik} - x_{Nik})\beta) + \exp(\alpha_i + (x_{Bik} - x_{Nik})\beta) + 1} \right]^{y_{Aik}} \left[ \frac{\exp(\alpha_i + x_{ik}\beta)}{\exp(\alpha_i + (x_{Aik} - x_{Nik})\beta) + \exp(\alpha_i + (x_{Bik} - x_{Nik})\beta) + 1} \right]^{y_{Bik}} \left[ \frac{1}{\exp(\alpha_i + (x_{Aik} - x_{Nik})\beta) + \exp(\alpha_i + (x_{Bik} - x_{Nik})\beta) + 1} \right]^{y_{Nik}} \quad (29)$$

Given the fundamental nonlinearity of the model, we cannot just use differences from within-group means (as we might do in a least-squares context) to sweep out the “intercept” values in the logit “index”, $\alpha_i + (x_{jik} - x_{Nik})\beta$. Instead we use a clever insight from Chamberlain. His approach relies on the sequences of choices observed in the set of choices for each person. Suppose there are just two choices for each person, as in equation (28) (multiple-alternatives and several choice occasions just mean messier algebra). Then the person could choose (1,1), (1,0), (0,1) or (0,0). Chamberlain conditioned the probability of a particular pattern of choices on the outcome that the sum of the indicators took on each particular value.
The conditional likelihood is given by:

$$L^* = \prod_{i=1}^{n} \Pr \left( Y_{i1} = y_{i1}, Y_{i2} = y_{i2} \left| \sum_{k=1}^{K} y_{ik} \right. \right)$$

(30)

For example, the probability that the pair of choices will be (0,1) when the sum of the indicators is one is given by:

$$\Pr(0,1| \text{sum} = 1) = \frac{\Pr(0,1)}{\Pr(0,1) + \Pr(1,0)}$$

(31)

The probability can be built from the binary probit probabilities in each of the two choice occasions. To simplify the notation, assume this is a case where the levels of the attributes have been normed on the status quo alternative, so that \( \alpha_i + (x_{ik} - x_{0k}) \beta \) can be written simply as \( \alpha_i + x_{ik} \beta \) with the regressors understood to be the difference in attribute levels between the “1” and the “0” alternatives. The error terms are still assumed to be uncorrelated, so the key insight is that each of the joint probabilities for the pairs of outcomes in the numerator and denominator of equation (31) can be written as the product of the probabilities of each outcome by itself:

$$\Pr(0,1| \text{sum} = 1) = \frac{1}{1 + \exp(\alpha_i + x_{i2} \beta)} \cdot \frac{\exp(\alpha_i + x_{i2} \beta)}{1 + \exp(\alpha_i + x_{i2} \beta)}$$

(32)

We can now see why this conditional likelihood is attractive...the denominator terms in the expressions above and below the line will cancel, leaving just:

$$\Pr(0,1| \text{sum} = 1) = \frac{\exp(\alpha_i + x_{i2} \beta)}{\exp(\alpha_i + x_{i2} \beta) + \exp(\alpha_i + x_{i1} \beta)}$$

(33)

The other way to get a sum of 1 will have a probability that is just the complement of this probability, with \( \exp(x_{i1} \beta) \) in the numerator instead.

In the case of two binary choices, there are just three possible sums: one way to get a sum of 2; two ways to get a sum of 1, and one way to get a sum of 0. Notice that somebody who chooses “all zeros” or “all ones” will yield a sum of zero or a sum of \( K \) (the number of choices, here just two). Since there is only one way to do each of these things, people who always choose...
the same alternative will have a conditional probability of one, and the log of one is zero, so they add nothing to the log of the conditional likelihood. Their choices will not contribute to the estimation of the slope parameters in the vector $\beta$. Only the cases with sums between 2 and $K-1$ are helpful.

When the objective function is constructed from these types of conditional probabilities, we allow for individual-specific “lumps” of utility in the amount of $\alpha_i$ for each respondent, although we forgo the ability actually to estimate these parameters (as we do in a fixed effects model in a least-squares context when the slope coefficients are estimated using the method of deviations from within-group means). However, the slope coefficients, $\beta$ (here interpreted as marginal utility parameters associated with each attribute) are estimated assuming the existence of heterogeneity in the $\alpha$ parameters.

### 4.3 Hausman test for fixed effects

To test a fixed effects logit against an ordinary logit, we normally use a Hausman-type test concerning what happens to the vector of slope coefficients. If preferences are homogeneous (i.e. if there is no need for the fixed effects model) both the ML and the CML are consistent, but the Chamberlain estimator is inefficient (because it will not really use the information from people who chose the same alternative on all of their choice occasions, and it does not take advantage of the constraint that $\alpha_i = \alpha$). The Hausman test is:

$$
\left( \hat{\beta}_{CML} - \hat{\beta}_{ML} \right) \left( Var[CML] - Var[ML] \right)^{-1} \left( \hat{\beta}_{CML} - \hat{\beta}_{ML} \right) \sim \chi^2(k)
$$

where $k$ is the number of slope parameters (e.g. marginal utilities in an additive RUM model). A large value of this test statistic says that moving to a fixed effects model has made a big enough difference in the slopes for us to believe that the homogeneous model was too restrictive.

Fixed effects logit models can be invoked in Stata by using the command:

```
clogit best x1 x2 ..., group(personid);
```

where the data have been entered with one row for each alternative, three rows for each choice set, and $k$ choice sets (typically five) per person. The variable “best” is a binary indicator for the chosen alternative in each choice set, and the majority of respondents in the sample will each account for fifteen rows ($5 \times 3$) in the data. The $x$ variables are the explanatory variables, both individual- and alternative-specific, which we use to account for respondents’ choices.

For our preferred specification, the results of the Hausman test are shown below. Notice that the differences in the estimated parameters are relatively minor and that the calculated $\chi^2$ test value rejects the null hypothesis only at the 13% level, although the algorithm is hampered by the fact that the difference between the parameter variance-covariance matrices for the two models is not positive definite (where the difficulty concerns the term in discounted recovered/remission years, a variable which is individually statistically significantly different from zero only at the 10% level in both the fixed effects model and the non-fixed effects model).

There is little a priori reason to anticipate that a fixed effects specification will be necessary because the levels of all of the main regressors have been assigned randomly across choice sets and across individuals. The only source of concern will stem from the appearance of the interaction terms in age and age-squared which shift three of the basic coefficients, and the
selection correction interaction term involving each individual’s fitted probability of participating in the estimating sample (relative to the original 525,000 recruiting contacts for Knowledge Networks). While the results of the Hausman test, below, suggest that there is no strong evidence of the need for a fixed effects conditional logit model, we employ it for its greater generality. Failure to exploit the panel dimension of our data would invite the criticism that we have somehow obscured relevant heterogeneity in consumer preferences.

```
hausman fixed nofixed;
```

Note: the rank of the differenced variance matrix (9) does not equal the number of coefficients being tested (13); be sure this is what you expect, or there may be problems computing the test. Examine the output of your estimators for anything unexpected and possibly consider scaling your variables so that the coefficients are on a similar scale.

```
---- Coefficients ----
|      (b)          (B)            (b-B)     sqrt(diag(V_b-V_B)) |
|     fixed       nofixed       Difference          S.E._
-------------+----------------------------------------------------------------
 b7term |    .0138705     .0136397        .0002308        .0001874
 dilog |   -49.15848    -48.05254       -1.105939        .8740273
 drlog |   -16.75242    -17.09082        .3383987               .
 dllog |   -561.9437    -500.5019       -61.44189        43.40268
 dllog_agenow |    19.63725     18.28144        1.355803        1.790943
 dllog_age~w2 |   -.1800343    -.1764901       -.0035442        .0174983
 dllog2 |   194.5601     175.4035        19.15658        11.55654
 dllog2_~w |   -7.504353    -7.121171       -.3831828        .5158851
 dllog2_age~w |   .0714078     .0710271        .0003806        .0053307
 dddllog |   -4.500657    -4.335806       -.1648511        .1710725
 dddllog ~w2 |   .0561213     .0545688        .0005525        .0020757
 sdilog |    3.372028     3.006647        .3653809        .4531743
------------------------------------------------------------------------------
 b = consistent under Ho and Ha; obtained from clogit
 B = inconsistent under Ha, efficient under Ho; obtained from clogit

Test: Ho: difference in coefficients not systematic
    chi2(9) =  (b-B)’[(V_b-V_B)^(-1)](b-B)
             =       13.77
    Prob>chi2 =  0.1307
(V_b-V_B is not positive definite)
```

5 Random-parameters logit models

Over the last decade, it has become increasingly easy to consider random-parameters variants of multiple discrete choice models. The familiar “mixed logit” model (e.g. Greene, 2008, p. 851-859) is an important alternative specification to consider in this application. Using Kenneth Train’s mxlmsl.m Matlab algorithm (mixed logit by maximum simulated likelihood), we have estimated mixed logit models assuming normal distributions for all thirteen
coefficients featured in Model 5 in Table 3 in the paper, with the following results.

A key insight concerns the distribution of ages in the estimating sample. This distribution is depicted in the figure. If we allow for normally distributed coefficients, all of the non-random variation due to age in Model 5 is absorbed instead by the random parameters. Furthermore, only the four basic parameters, $\beta_0$ (on the net income term), and $\alpha_{10}$, $\alpha_{20}$, and $\alpha_{30}$ (on each of the discounted future health state terms) display statistically significant heterogeneity (in terms of the estimated dispersion in the random parameter).

### 5.1 Results: Random parameters specifications

If we generalize Model 5 to allow for interaction terms in age and age-squared to shift the estimated $\beta_0$ parameter, neither of these interaction terms bears a statistically significant coefficient, so there is apparently some other source of unobserved heterogeneity in this marginal utility of income parameter. However, it seems clear that the age variable is a prominent contributor to heterogeneity in the slope coefficients which capture the marginal (dis)utility of future adverse health states. This heterogeneity with respect to age is a key consideration in any model of health risk reduction preferences, so we expressly do not wish to subsume it with all other unspecified sources of heterogeneity in a mixed logit model. Thus we opt for a conventional non-random parameters specification in this application. In other research using these data, we explore for other possible dimensions of heterogeneity in the marginal utility of income parameter, but those analyses are beyond the scope of the main paper (Cameron and DeShazo, 2009).

### 6 Alternate Specifications

#### 6.1 Preliminary models

In Table 2, we consider first the implications of our data in the context of the simplest ad hoc specification. Model E1 reveals that the two main features of each program we describe in our choice scenarios—namely, its cost and the size of the risk reduction it would achieve—are both strongly statistically significant determinants of people’s choices. We then show, in Model E2, that the two most important features of each illness profile—namely, the prospective sick-years and lost life-years—are also strongly statistically significant in explaining choices. In the final specification in this table, Model 3, we implement the four-parameter structural model in equation ( ), imposing a Box-Cox transformation with $\lambda = 0.42$ as the function $f(\cdot)$.2

Model E3 is a homogenous-preferences specification, estimated without sign restrictions, and shows robust significance and the expected signs on all four primary parameters. The estimated marginal utility of income is positive and declines with the level of income. The marginal utilities of discounted prospective sick-years, post-illness recovered/remission-years, and lost life-years are all negative and very strongly significantly different from zero. Simple intuition might suggest that death should be perceived as far “worse” than illness and recovery/remission. However, it is important to keep in mind that the units involved are

---

2 The curvature in the net income term allows for risk aversion with respect to financial risk. Eeckhoudt and Hammit (2004) find that this type of risk aversion increases WTP for risk reductions in definable cases, but that in general, the relationship is theoretically ambiguous. We note that the structural form in Model 3 yields a somewhat poor overall fit than that attained with ad hoc Model 2. However, this structural form is the feature that permits us to calculate rigorously the corresponding option price that is our WTP measure.
discounted single years in each health state. In many illness profiles, there are more life-years lost than there are sick-years, but the lost life-years are always further into the future, so they are discounted more heavily. Thus the marginal utility per discounted health-state year does not convey the overall disutility of total future time in that state. Also, the relatively large (dis)utility associated with recovered/remission state reflects the seriousness of the major illnesses our survey describes. Rightfully, respondents do not interpret being recovered or in remission from any of this list of major illnesses as being equivalent to the pre-illness “healthy” state, which would produce a zero coefficient. For example, there may be considerable anticipated disutility from the prospect of living as a cancer or heart-attack survivor, relative to the respondent’s current health.3

In the main paper, Cameron and DeShazo (2009), we quickly relax the assumption that the marginal utilities from each prospective future health state are independent of the duration of that state and the durations of other health states that characterize the illness profile in question.

6.2 Appropriate transformation for health state durations
In the main paper, Cameron and DeShazo (2009) we first consider a model that is linear in the discounted prospective sick-years, recovered/remission years, and lost life-years. The parameters of this model are identical to the underlying parameters in the future-period indirect utility function. However, we find that a model which takes the present discounted time in each future health state as characteristics of the entire illness profile may be superior, and that a shifted logarithmic transformation seems to dominate an ordinary linear function of the present discounted durations.

It is impractical to do a four-way grid search to establish different transformations for each of the four main variables in the estimating specification. However we have constrained the Box-Cox parameter for the net income term, \( f(Y) \) to be \( \lambda = 0.42 \) and we have conducted a single line-search across values of \( \theta \), a Box-Cox transformation parameter shared by all of the discounted durations in each health state (also shifted by one, to ensure that a zero duration corresponds to a zero value of the transformed variable as well). In Figure E2, we show that the maximized log-likelihood for the four-parameter conditional-logit model is relatively insensitive to the choice of \( \theta \) for values less than zero (which would correspond to the logarithmic transformation adopted in the paper). However, the maximized log-likelihood begins to drop off at a distinctly faster rate for parameter values greater than zero.

Based on Figure E2, we elect to use the simple shifted logarithmic transformation for each of the discounted health-state durations in our model.

6.3 Different assumptions about the individual discount rate
In addition to the specifications reported in the main paper, Cameron and DeShazo (2009), we have explored a variety of other possible specifications. One key assumption in the estimation concerns the common discount rate attributed to all respondents.

The results in the main paper reflect the assumption of a 5% discount rate. However, we also report the distributions of simulated WTP estimates if alternative assumptions are made about this discount rate. These alternative WTP results are derived using the parameter estimates reported in Table E1 in this appendix. The discount rate assumption is invoked during the construction of variables used in our estimation. When a different discount rate is assumed,

---

3 The evidence about the marginal (dis)utility of a discounted recovered/remission-year also do not involve diabetes or Alzheimer’s disease, since it was not possible to describe credible scenarios with recovery from these diseases.
different “present discounted” variables must be calculated based on that discount rate. Since these variables will be somewhat different, so will be all of the parameter estimates produced by the model.

Table E2 reports the estimated “turning points” of the three main utility parameters which are specified as quadratic functions of the respondent’s current age. These turning points are maximum or minimum value of the marginal utility, as indicated. The overall age profile of willingness to pay reflects the combined effects of all three of these quadratic forms, acting differently on the three utility parameters. Our models are not simply ad hoc specifications where \( WTP \) is allowed to depend directly on age. The age effects are mediated by our structural model of preferences.

6.4 Including an alternative-specific dummy for “either program”

There is no natural ordering to the risk reduction programs offered in each choice set, since ten illness labels are randomly selected from a possible eleven illnesses or injuries for each gender. These ten illnesses are randomly paired into our three-alternative choices sets (in addition to the “Neither Program” alternative). The order in which the ten illnesses appear for any individual is thus random. Consequently, there is no real argument for alternative specific dummy variables on the “left” and “right” alternatives in the substantive pair.

However, researchers are sometimes interested in knowing whether there is some unobserved bias either for or against either of the substantive alternatives (versus the status quo “Neither Program” alternative). Testing for such an effect can be done either with a “status quo” dummy variable, or an “Either Program” dummy associated with both of the offered programs. We use the latter option in a model presented in the second column of results in Table E3. (The first column reproduces the results for Model 2 in the main paper.)

Clearly, allowing for there to be some unspecified “lump” of utility associated with either of the two risk reduction programs produces a strongly significant positive point estimate on the additional dummy variable, as well as a large increase in the maximized value of the log-likelihood. We infer that respondents are somewhat inclined to choose one of the two risk reduction programs regardless of the costs and benefits of either program.

A variable such as this is typically employed to capture the net effects of phenomena such as “yea-saying” (which would tend to produce a positive coefficient on this variable) or “payment vehicle rejection” (which would tend to produce a negative coefficient). In our data, therefore, it seems that there is some autonomous utility derived from either risk-reduction alternative, but not from the status quo. This could be what stated-preference researchers sometimes call “warm glow.” It could be that the benefits of the risk-reduction programs are perceived to be greater than we describe by some amount unrelated to the quantitative attributes described in our choice scenarios, or perhaps the program costs are perceived to be lower than the scenarios state. Unfortunately, as is always the case with these variables, it is impossible to know exactly what this type of variable is capturing.

The consequences of including this dummy variable, and then simulating it to have a value of zero during our \( WTP \) calculations, is shown in the second column of results in Table E4. As expected, canceling out this positive lump of utility shared by all risk-reduction programs leads to a modest decrease in \( WTP \) for each of our basic set of five illness profiles. The first column again reproduces the results for the final model in the main paper. For example, \( WTP \) to avoid sudden death in the current period drops from $5.82 to $4.55.

The logic for netting out the autonomous portion of \( WTP \) captured by the “either program” dummy variable, however, is not clear. When the estimated coefficient is negative, and
there are reasons to suspect that there is a significant problem with payment vehicle rejection (as in the use of taxes to pay for public goods), it may be defensible to cancel the autonomous negative component in \( WTP \). In this case, however, respondents appear to be willing to pay some amount for any type of health risk reduction program, regardless of the size of its effect or the type of the risk. Perhaps it is not unreasonable that people might be willing to do this. None of our costly programs yields a zero risk reduction, so we cannot test for a positive \( WTP \) even when benefits are zero. Perhaps \( WTP \) is not strictly proportional to the size of the risk reduction, and this is what the “either program” dummy variable is picking up.

6.5 If respondents anticipate having only half as much income when sick
In the main model in the paper, we make the assumption that respondents do not anticipate having a substantially reduced income, should they suffer the illness or injury described in each illness profile. However, it is straightforward to adjust the calculation of the variables for use in the main model to accommodate other assumptions about income levels during illness years.

If the respondent expects to earn only \( \gamma Y \) during any years when he or she is suffering from a major illness, then the formula in expression (7) must be adapted. It becomes

\[
\begin{align*}
&f(Y - c)[\Pi^{AH} pdvc + \Pi^{AS} pdvp] \\
&- f(Y)[\Pi^{NH} pdvc + \Pi^{NS} pdvp] + f(\gamma Y)[\Delta \Pi^{AS} pdvi]
\end{align*}
\]

The third model in Table E3, and the third set of \( WTP \) simulations in Table E4 provide the details concerning the effects of this adjustment on our estimates. The impact of this change is very small. This is because for many people, the discounted number of sick-years is relatively small compared to the discounted remaining lifespan.

6.6 If respondents perceive other costs in addition to those quoted
One reviewer of the main paper was concerned that some respondents may have treated the stated costs of each program as less than the full opportunity cost that would be involved if they chose to participate. On Form 17, we state specifically that the risk reduction programs in question would not involve “uncomfortable procedures.” We do state that “Your participation in a program would cost you money.” These programs would not be covered by the respondent’s current health insurance. “These higher costs might take the form of a co-payment when you visit your doctor or higher monthly health insurance costs.” “To make it easier to compare, we present all costs as monthly costs, and also as annual costs. You would need to pay for, and participate in, a program for the next ___ years to get its benefits.” (The precise number of years corresponded to that individual’s current age and nominal gender-specific life expectancy.) We did not explicitly limit the cost of the program to simply the cost of the test. Instead, we were careful to refer to the “cost of the program” (where the programs are described on Form 17 as involving prescribed “medication and life-style changes that reduce your risk of getting the illness”).

Earlier in the survey, however, on Form 7, we specifically asked respondents to consider the difficulty of making life-style changes. We asked them: “Changing your lifestyle or habits can be difficult because it requires time, money, and effort. How difficult would it be for you to do the following things?” The listed options included the following measures: drink less alcohol, quit smoking, eat a healthier diet, see a doctor more regularly, exercise more, lose weight, use a
seatbelt more. We went through one phase of survey development with language in the instrument where we tried to explain the idea of the monetized disutility of the tests themselves, and opportunity costs and the full cost of time. However, without getting into discussions of the value of travel time to the doctor’s office and the pharmacy, and the prospective disutility of a new exercise regimen or dietary restrictions, there seemed to be no happy medium, so we opted for a minimalist approach. Perhaps there would have been a better option, but we could not see it at the time. To meet the length/duration restrictions under our contract with Knowledge Networks, of course, it was necessary to prune many things out of the survey that we were keenly interested to include. This is a frequent problem with survey research in general.

In response to this concern, however, we have investigated additional models where we allow the estimated marginal utility of net income to depend on the respondent’s answers to our questions about the difficulty of accomplishing lifestyle changes. We take advantage of the wording on Form 7 in the question: “Changing your lifestyle or habits can be difficult because it requires time, money, and effort. How difficult would it be for you to do the following things?” A slight complication is that respondents were only asked about each of these things if they responded on Form 6 that there was still at least some room for them to reduce their health risks by improving their lifestyle or habits in these ways. We assume that if the individual reports no room to improve along any particular dimension, then it would be very hard at the margin for them to improve any further on this dimension. (Cleaning up a few of your bad habits may be relatively easy, but getting rid of all of them might be tough.)

However, if there is still room to improve on one or more dimensions, and respondents report that it would be easy or difficult for them to do so, this is the notion we wish to capture. We construct a crude variable to measure “ease of improving health habits.” For each type of the seven health habits identified on Form 6 and Form 7, we build two variables. One is prefixed by “improve_” and measures “opportunity for improvement” with ratings that vary from 0 = “no opportunity for improvement” to 4 = “much room to improve.” The second variable is prefixed by “easy_” and measures the ease with which these available improvements in health habits could be accomplished. For this variable, we have inverted the question about how difficult it would be to make improvements. For our “easy_” variables, the ratings are coded as 0 = “hard to improve” to 4 = “easy to improve.”

For each of the seven health habits, we construct an interaction between the “improve_” and “easy_” variables. This interaction term is zero if the individual has no opportunity to improve or if they do, but it would be very hard for them to do so. This interaction term takes on a larger value (to a maximum value of 16) if there is lots of room for the individual to improve their health habits and they believe it would be easy to do so. Acknowledging the degree of approximation involved in the use of ratings, and the different metrics across the different questions, we then forge ahead and add these interacted ratings across all seven types of health habits to generate a variable that may serve as a proxy for the likely psychic or non-pecuniary costs to the individual if they need to make “lifestyle changes” in addition to paying for the annual pin-prick blood test in the choice scenarios.

The maximum value for our constructed indicator is 16x7 = 112. It measures “ease of making lifestyle changes.” We desire a variable that will be larger if the implicit costs to the individual of making these changes is larger, so we subtract our indicator from 112 to convert it into an indicator called hard, which proxies for the “difficulty of making lifestyle changes.” As a further complication, however, not all respondents answered all of the questions on Form 6 and Form 7, so we create an indicator for whether information was missing. 1,724 of our 1,801
respondents provided sufficient information to build this variable. We thus use a second indicator variable to control for data availability.

Now we simplify the intuition by supposing that the indirect utility difference that drives program choices is linear in net income and we don’t need to worry about the pattern of net income across the uncertain prospects of getting sick or remaining healthy. In that simple case, $\beta (Y - c) - \beta (Y) = \beta (-c)$. Suppose costs are perceived as systematically higher than what is stated in the choice scenario, say $c\theta$, where $\theta > 1$. If respondents are reacting to this larger cost, but we control only for $c$, then we will actually be estimating $(\beta \theta)(-c)$, and the apparent “marginal utility of net income” coefficient will be too large. This coefficient forms the denominator of the WTP function, so a too-large value will lead to a WTP estimate that is too small. People who look like they are unwilling to pay the amount stated in the choice scenario are actually unwilling to pay the larger amount. Failure to accommodate these other implicit costs will lead to underestimates of WTP.

We incorporate our new variable, $hard$, along with the indicator for its availability, into our model by allowing these two variables to shift the $\beta$ coefficient. The slope coefficient on the interaction with the indicator variable is insignificant, but the slope coefficient on the interaction with $hard$ is positive and strongly significant. If we estimate $\beta$ as a scalar, its point estimate is 0.0139. We do not constrain the systematically varying version of this parameter to be positive, so a few negative values result. The figure shows the range of implied values for the $\beta$ parameter in this more-general model. The mean of these fitted values is 0.0145 and the median is 0.0154.

Thus there exist a range of perceived difficulties of making life-style changes among our respondents. The values of the $hard$ variable range from 0 through 112, with a median of 92 and an interquartile range of 82 through 99. For people who perceive life-style changes as relatively more difficult (i.e. those who may consider other implicit costs associate with each risk reduction program), the marginal utility of income is estimated to be higher, which would imply a lower WTP for the risk reduction programs in the choice scenarios. For people who perceive life-style changes as relatively easier, the marginal utility of income is estimated to be lower, which would imply a higher WTP for the risk-reduction programs in the choice scenarios.
As an alternative, we could build the $hard_{i}$ variable using only the information on how easy it would be to improve each health-related behavior on the list. When we do this, the implied $\beta$ parameters display the range shown in this second histogram. In this case, there are fewer negative fitted values, but the results are qualitatively the same. The slope coefficient on the interaction between the $hard_{i}$ variable and the net income term is positive and strongly statistically significant.

The relevant question, now, is “what would people have been willing to pay had they believed that the quoted cost on the survey was the full cost of the program—i.e. that there were no additional costs associated with the difficulty of complying with the lifestyle changes that might be required?” It might be tempting to simulate the value of the marginal utility of income parameter for the case where everyone believes that it is trivially easy to implement life-style changes. This would correspond to the counterfactual where nobody perceives any implicit costs of this variety in addition to the cost of having the test.

We had intended to do this sort of thing in our analysis, which was why we collected the information on Form 6 and Form 7. However, we did not anticipate that respondents might view “lifestyle changes” in two separate ways. We expected that people would view them as necessary complements to the health testing programs described in the choice scenarios. This is the implicit assumption behind the concern that respondents will impute other costs to each program besides just the cost stated in the survey question. However, it may actually be the case that respondents view the testing programs in the survey as substitutes for the lifestyle changes that they know they should really be trying to make. If they perceive that participation in these testing programs will allow them the luxury to continue with their current poor health habits but still lower their health risks, they may actually express greater demand—because the perceived benefits are greater than just the reduction of health risks.

This makes things considerably more complicated. If we were to simulate a situation where everyone found it perfectly easy to implement any required life-style changes that would be required along with the testing program, the marginal utility parameter for income would be vastly smaller, causing the inferred WTP for these programs to be vastly bigger. But here’s the catch: if life-style changes were easy, the “price of a substitute” for the testing program would also be dramatically smaller, which would decrease the demand for the testing programs. People could simply change their health habits and they would have no need for the testing program. Thus it seems highly inappropriate to consider any adjustments to the stated cost of the program without making corresponding adjustment to the price of substitutes. Clearly, more research is needed, and it should focus on this “complements versus substitutes” distinction.
Incidentally, we do have some evidence, in other work with these data, for the “substitutes” possibility. In our research concerning the disease labels, non-smokers are willing to pay very little for tests to reduce their risk of lung cancer or respiratory disease, whereas smokers are willing to pay amounts for these two illnesses that substantially exceed the WTP amounts measured for all other illnesses for the general population. In this case, the substitution effect appears to dominate very strongly.
### 7 Tables

#### 7.1 Table E1 – Preliminary simple models

<table>
<thead>
<tr>
<th>Table E1 – Ad Hoc Models Versus Simplest Structural Model</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>(Individuals = 1,801, Completed Choice Sets = 7,520; No Selection Correction, Fixed Effects Conditional Logit Estimates)</strong></td>
</tr>
<tr>
<td>Model</td>
</tr>
<tr>
<td>-------</td>
</tr>
<tr>
<td>Model</td>
</tr>
<tr>
<td>Monthly cost of program</td>
</tr>
<tr>
<td>Risk reduction: $</td>
</tr>
<tr>
<td>Sick-years</td>
</tr>
<tr>
<td>Lost life-years</td>
</tr>
<tr>
<td>Maximized log-likelihood</td>
</tr>
</tbody>
</table>

---

*a* Each respondent is asked to consider five choice sets, so these are panel data. We use the maximum likelihood estimator that biostatisticians and epidemiologists call “conditional logistic regression for matched case-control groups” and that economists and other social scientists call “fixed-effects logit for panel data.” The estimator is coded as “clogit” in the Stata software package. See Greene (2008, p. 800-806).

*b* Absolute asymptotic t-test statistics in parentheses (**=statistically significant at the 1% level; **=statistically significant at the 5% level).

*c* The superscript in parentheses denotes a Box-Cox transformation with the indicated parameter value:

$$X^{(\lambda)} = \left( \frac{X}{\lambda} - 1 \right) / \lambda.$$  

The value of 0.42 for $\lambda$ was determined by a line-search in our more detailed models. Standard errors are of course conditional on this value for $\lambda$. We have previously used square root or quadratic transformations as approximations that dominated either a linear or a logarithmic function for the net income variable, but any of these transformations produces a very strongly statistically significant coefficient.
### Table E1 – Different assumptions about the discount rate

**TABLE E1 – EFFECT OF DISCOUNT RATE ASSUMPTION ON PARAMETER ESTIMATES**

(1,801 INDIVIDUALS, 7,520 COMPLETED CHOICE SETS, 22,560 ALTERNATIVES)

<table>
<thead>
<tr>
<th>(Parameter) Variable</th>
<th>Model 5</th>
<th>Model 5</th>
<th>Model 5</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>r=.03</td>
<td>r=.05</td>
<td>r=.07</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_0 \left( (Y_i - c_i^j)^{(42)} cterm_i^j - (Y_i)^{(42)} yterm_i^j \right)$</td>
<td>0.01107</td>
<td>0.01387</td>
<td>0.01644</td>
</tr>
<tr>
<td></td>
<td>(9.43)***</td>
<td>(9.44)***</td>
<td>(9.35)***</td>
</tr>
<tr>
<td>$(\alpha_{10}) \Delta \Pi_i^{JS} \log (pdvi_i^j + 1)$</td>
<td>-41.01</td>
<td>-49.16</td>
<td>-57.38</td>
</tr>
<tr>
<td></td>
<td>(5.16)***</td>
<td>(5.64)***</td>
<td>(5.99)***</td>
</tr>
<tr>
<td>$(\alpha_{13}) [P(sel_i) - P^*] \Delta \Pi_i^{JS} \left[ \log (pdvi_i^j + 1) \right]$</td>
<td>2.777</td>
<td>3.378</td>
<td>4.003</td>
</tr>
<tr>
<td></td>
<td>(2.32)**</td>
<td>(2.35)**</td>
<td>(2.33)**</td>
</tr>
<tr>
<td>$(\alpha_{20}) \Delta \Pi_i^{JS} \log (pdvr_i^j + 1)$</td>
<td>-13.11</td>
<td>-16.75</td>
<td>-20.84</td>
</tr>
<tr>
<td></td>
<td>(1.75)*</td>
<td>(1.79)*</td>
<td>(1.78)*</td>
</tr>
<tr>
<td>$(\alpha_{30}) \Delta \Pi_i^{JS} \log (pdvl_i^j + 1)$</td>
<td>-344.7</td>
<td>-561.9</td>
<td>-862.9</td>
</tr>
<tr>
<td></td>
<td>(2.30)**</td>
<td>(3.15)***</td>
<td>(3.94)***</td>
</tr>
<tr>
<td>$(\alpha_{31}) age_{t0} \Delta \Pi_i^{JS} \log (pdvl_i^j + 1)$</td>
<td>12.44</td>
<td>19.64</td>
<td>29.51</td>
</tr>
<tr>
<td></td>
<td>(2.03)**</td>
<td>(2.71)***</td>
<td>(3.37)***</td>
</tr>
<tr>
<td>$(\alpha_{32}) age_{t0}^2 \Delta \Pi_i^{JS} \log (pdvl_i^j + 1)$</td>
<td>-1.198</td>
<td>-1.8</td>
<td>-2.606</td>
</tr>
<tr>
<td></td>
<td>(2.03)**</td>
<td>(2.59)***</td>
<td>(3.15)***</td>
</tr>
<tr>
<td>$(\alpha_{40}) \Delta \Pi_i^{JS} \left[ \log (pdvl_i^j + 1) \right]^2$</td>
<td>93.92</td>
<td>194.6</td>
<td>372.1</td>
</tr>
<tr>
<td></td>
<td>(1.60)</td>
<td>(2.36)**</td>
<td>(3.18)***</td>
</tr>
<tr>
<td>$(\alpha_{41}) age_{t0} \Delta \Pi_i^{JS} \left[ \log (pdvl_i^j + 1) \right]^2$</td>
<td>-3.941</td>
<td>-7.504</td>
<td>-13.63</td>
</tr>
<tr>
<td></td>
<td>(1.61)</td>
<td>(2.23)**</td>
<td>(2.94)***</td>
</tr>
<tr>
<td>$(\alpha_{42}) age_{t0}^2 \Delta \Pi_i^{JS} \left[ \log (pdvl_i^j + 1) \right]^2$</td>
<td>0.04036</td>
<td>0.07141</td>
<td>0.123</td>
</tr>
<tr>
<td></td>
<td>(1.67)*</td>
<td>(2.20)***</td>
<td>(2.80)***</td>
</tr>
<tr>
<td>$(\alpha_{50}) \Delta \Pi_i^{JS} \left[ \log (pdvi_i^j + 1) \right] \times \left[ \log (pdvl_i^j + 1) \right]$</td>
<td>75.48</td>
<td>104</td>
<td>143.4</td>
</tr>
<tr>
<td></td>
<td>(1.60)</td>
<td>(1.42)</td>
<td>(1.31)</td>
</tr>
<tr>
<td>$(\alpha_{51}) age_{t0} \Delta \Pi_i^{JS} \left[ \log (pdvi_i^j + 1) \right] \times \left[ \log (pdvl_i^j + 1) \right]$</td>
<td>-3.401</td>
<td>-4.501</td>
<td>-5.926</td>
</tr>
<tr>
<td></td>
<td>(1.79)*</td>
<td>(1.58)</td>
<td>(1.42)</td>
</tr>
<tr>
<td>$(\alpha_{52}) age_{t0}^2 \Delta \Pi_i^{JS} \left[ \log (pdvi_i^j + 1) \right] \times \left[ \log (pdvl_i^j + 1) \right]$</td>
<td>0.04304</td>
<td>0.05612</td>
<td>0.07237</td>
</tr>
<tr>
<td></td>
<td>(2.34)**</td>
<td>(2.10)**</td>
<td>(1.89)*</td>
</tr>
<tr>
<td><strong>Control for sample selection</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$(\alpha_{13}) [P(sel_i) - P^*] \Delta \Pi_i^{JS} \left[ \log (pdvi_i^j + 1) \right]$</td>
<td>2.777</td>
<td>3.378</td>
<td>4.003</td>
</tr>
<tr>
<td></td>
<td>(2.32)**</td>
<td>(2.35)**</td>
<td>(2.33)**</td>
</tr>
<tr>
<td>Log L</td>
<td>-11684.9</td>
<td>-11685.1</td>
<td>-11687.5</td>
</tr>
</tbody>
</table>
### Table E2 – Turning points in age profiles of utility parameters

(For the three models in Table E1)

| Age at max. of 1\textsuperscript{st} age profile: on $\log(\text{pdvl}_i^j + 1)$ | 51.9 | 54.5 | 56.6 |
| Age at min. of 2\textsuperscript{nd} age profile: on $\left[\log(\text{pdvl}_i^j + 1)\right]^2$ | 48.8 | 52.5 | 55.4 |
| Age at min. of 3\textsuperscript{rd} age profile: on $\left[\log(\text{pdvl}_i^j + 1)\right] \times \left[\log(\text{pdvl}_i^j + 1)\right]$ | 39.5 | 40.1 | 40.9 |
### 7.4 Table E3 – “Either program” model and “half-income while sick” models

**Table E3—Translog-type fixed effects conditional logit models; Two alternative specifications**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Constructed Variable</th>
<th>Model 2 in main paper</th>
<th>“Either program” dummy</th>
<th>Half-income while sick</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_0$</td>
<td>$(Y_i - c_i^j)^{(0.42)}$</td>
<td>0.0139</td>
<td>0.0153</td>
<td>0.01384</td>
</tr>
<tr>
<td>$\alpha_{10}$</td>
<td>$\Delta \Pi_i^{\text{IS}} \log (pdvi_i^j + 1)$</td>
<td>-49.2</td>
<td>-35.3</td>
<td>-46.59</td>
</tr>
<tr>
<td>$\alpha_{20}$</td>
<td>$\Delta \Pi_i^{\text{IS}} \log (pdvr_i^j + 1)$</td>
<td>-16.8</td>
<td>-13.5</td>
<td>-17.14</td>
</tr>
<tr>
<td>$\alpha_{30}$</td>
<td>$\Delta \Pi_i^{\text{IS}} \log (pdvl_i^j + 1)$</td>
<td>-562.4</td>
<td>-512.</td>
<td>-562.4</td>
</tr>
<tr>
<td>$\alpha_{31}$</td>
<td>$\ldots \text{age}_{i0} \times \Delta \Pi_i^{\text{IS}} \log (pdvl_i^j + 1)$</td>
<td>19.6</td>
<td>19.0</td>
<td>19.62</td>
</tr>
<tr>
<td>$\alpha_{32}$</td>
<td>$\ldots \text{age}^2_{i0} \times \Delta \Pi_i^{\text{IS}} \log (pdvl_i^j + 1)$</td>
<td>-0.180</td>
<td>-0.175</td>
<td>-1.799</td>
</tr>
<tr>
<td>$\alpha_{40}$</td>
<td>$\Delta \Pi_i^{\text{IS}} \left[ \log (pdvl_i^j + 1) \right]^2$</td>
<td>195.2</td>
<td>177.</td>
<td>194.8</td>
</tr>
<tr>
<td>$\alpha_{41}$</td>
<td>$\ldots \text{age}_{i0} \times \Delta \Pi_i^{\text{IS}} \left[ \log (pdvl_i^j + 1) \right]^2$</td>
<td>-7.50</td>
<td>-7.17</td>
<td>-7.508</td>
</tr>
<tr>
<td>$\alpha_{42}$</td>
<td>$\ldots \text{age}^2_{i0} \times \Delta \Pi_i^{\text{IS}} \left[ \log (pdvl_i^j + 1) \right]^2$</td>
<td>0.0714</td>
<td>0.0684</td>
<td>0.07147</td>
</tr>
<tr>
<td>$\alpha_{50}$</td>
<td>$\Delta \Pi_i^{\text{IS}} \left[ \log (pdvi_i^j + 1) \right] \times \left[ \log (pdvl_i^j + 1) \right]$</td>
<td>104.3</td>
<td>93.0</td>
<td>102.9</td>
</tr>
<tr>
<td>$\alpha_{51}$</td>
<td>$\ldots \text{age}_{i0} \times \Delta \Pi_i^{\text{IS}} \left[ \log (pdvi_i^j + 1) \right] \times \left[ \log (pdvl_i^j + 1) \right]$</td>
<td>-4.50</td>
<td>-4.42</td>
<td>-4.454</td>
</tr>
<tr>
<td>$\alpha_{52}$</td>
<td>$\ldots \text{age}^2_{i0} \times \Delta \Pi_i^{\text{IS}} \left[ \log (pdvi_i^j + 1) \right] \times \left[ \log (pdvl_i^j + 1) \right]$</td>
<td>0.0561</td>
<td>0.0552</td>
<td>0.05563</td>
</tr>
</tbody>
</table>
Systematic selection correction term:

\[ \alpha_{13} \quad \ldots \left[ P(\text{sel}) - \bar{P} \right] \times \Delta \Pi_j^{s} \log \left( pdv_{ij} + 1 \right) \]

Alternative-specific dummy, “any program”

Maximized log-likelihood

<table>
<thead>
<tr>
<th></th>
<th>(2.10)**</th>
<th>(2.06)**</th>
<th>(2.08)**</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \alpha_{13} )</td>
<td>3.37</td>
<td>3.388</td>
<td>3.374</td>
</tr>
<tr>
<td></td>
<td>(2.34)**</td>
<td>(2.33)**</td>
<td>(2.35)**</td>
</tr>
<tr>
<td></td>
<td>0.200</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(4.76)***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-11685.12</td>
<td>-11673.79</td>
<td>-11684.81</td>
</tr>
</tbody>
</table>
7.5 **Table E4 – WTP estimates for “either program” and “half-income” models**

<table>
<thead>
<tr>
<th>Illness profile: age 45 now; …at 45:</th>
<th>r=0.05</th>
<th>“Either program” dummy</th>
<th>Half-income while sick</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Sudden death</td>
<td>$ 5.82</td>
<td>$ 4.55</td>
<td>$ 5.78</td>
</tr>
<tr>
<td></td>
<td>(3.78, 7.79)</td>
<td>(2.77, 6.48)</td>
<td>(3.79, 7.96)</td>
</tr>
<tr>
<td>2. 1 yr sick; nonfatal</td>
<td>2.96</td>
<td>2.00</td>
<td>2.92</td>
</tr>
<tr>
<td></td>
<td>(1.47, 4.64)</td>
<td>(0.65, 3.52)</td>
<td>(1.46, 4.58)</td>
</tr>
<tr>
<td>3. 5 yrs sick; nonfatal</td>
<td>4.47</td>
<td>2.99</td>
<td>4.35</td>
</tr>
<tr>
<td></td>
<td>(3.06, 6.18)</td>
<td>(1.62, 4.58)</td>
<td>(2.96, 5.96)</td>
</tr>
<tr>
<td>4. 1 yr sick; then die</td>
<td>5.76</td>
<td>4.81</td>
<td>5.66</td>
</tr>
<tr>
<td></td>
<td>(3.84, 7.75)</td>
<td>(3.12, 6.52)</td>
<td>(3.86, 7.8)</td>
</tr>
<tr>
<td>5. 5 yrs sick; then die</td>
<td>5.43</td>
<td>4.81</td>
<td>5.34</td>
</tr>
<tr>
<td></td>
<td>(3.53, 7.71)</td>
<td>(2.83, 6.80)</td>
<td>(3.32, 7.75)</td>
</tr>
</tbody>
</table>

*Notes: Units are in 2003 US dollars per microrisk reduction for each of five arbitrarily selected illness profiles (rows). Entries reflect 1000 random draws from the joint distribution of estimated parameters. We report the median, 5th and 95th percentiles for the sampling distribution of calculated WTP. Income is set at $42,000.*
8 Figures

8.1 Figure E1 – Three examples of illness profiles

Below are three illustrations of some of the range of illness profiles that we describe to respondents in the choice scenarios with which they are presented. Each scenario involves only a single spell of illness, rather than multiple spells, which were beyond the scope of the study. The individual’s current age and gender define the number of remaining nominal life-years at stake. The individual is informed that they face an existing baseline risk of each featured illness profile. The programs they are asked to consider do not change the time profile of the illness in question, only its likelihood of occurring.

The design of choice scenarios where interventions which are the subject of the choice exercise specifically alter the time profiles of the illnesses described is left to future research. Nevertheless, our fitted models can permit simulation of a change in the time profile of an illness—via a reduction in the probability of one illness profile, combined with a corresponding increase in the probability of a different illness profile.

**Illness Profile 1:** Sudden death in the current period (usual VSL illness profile)

<table>
<thead>
<tr>
<th>Illness Profile</th>
<th>Lost Life Years</th>
</tr>
</thead>
<tbody>
<tr>
<td>Health Status</td>
<td></td>
</tr>
</tbody>
</table>

**Illness Profile 2:** A nonfatal illness (with recovery) that reduces life expectancy

<table>
<thead>
<tr>
<th>Illness Profile</th>
<th>Latency Period</th>
<th>Sick Years</th>
<th>Recovered Years</th>
<th>Lost Life Years</th>
</tr>
</thead>
<tbody>
<tr>
<td>Health Status</td>
<td>healthy</td>
<td>sick</td>
<td>recovered</td>
<td></td>
</tr>
</tbody>
</table>

**Illness Profile 3:** A fatal illness (no recovery)

<table>
<thead>
<tr>
<th>Illness Profile</th>
<th>Latency Period</th>
<th>Sick Years</th>
<th>Lost Life Years</th>
</tr>
</thead>
<tbody>
<tr>
<td>Health Status</td>
<td>healthy</td>
<td>sick</td>
<td></td>
</tr>
</tbody>
</table>
8.2 Figure E2 – \( \log L \) as a function of health state duration transformation

Rationale for using shifted log transformation of each discounted prospective health state duration: consequences of line search across Box-Cox transformation parameters

8.3 Figure E3 – WTP, sudden death now, by discount rate

\( WTP \) for a microrisk reduction for sudden death now, as a function of respondent age now, for three different discount rate assumptions
8.4  Figure E4 – WTP, half-year sick, die half-year early, by discount rate

WTP for a microrisk reduction for six month reduction in life expectancy, preceded by six months of major illness, as a function of age now, for three different discount rate assumptions

8.5  Figure E5 – WTP, sudden death now, by income level

WTP for 1/1,000,000 reduction in risk of sudden death in the current period, as a function of respondent household income now in $’000, for a 45-year-old
9 References


Case-Control Studies, Vol 1. Lyon: IARC.

Cameron, T.A., J.R. DeShazo. 2009. “Demand for health risk reductions,” manuscript, 
Department of Economics, University of Oregon, Eugene, OR.

47: 225-238.


Wiley.