

# The value of health—Empirical issues when estimating the monetary value of a quality-adjusted life year based on well-being data

Sebastian Himmler<sup>1</sup>  | Jannis Stöckel<sup>1</sup>  | Job van Exel<sup>1,2</sup>  |  
Werner B. F. Brouwer<sup>1,2</sup> 

<sup>1</sup>Erasmus School of Health Policy & Management, Erasmus University Rotterdam, Netherlands

<sup>2</sup>Erasmus School of Economics, Erasmus University Rotterdam, Netherlands

## Correspondence

Sebastian Himmler and Jannis Stöckel, Erasmus School of Health Policy & Management, Erasmus University Rotterdam, Netherlands.

Email: [himmler@eshpm.eur.nl](mailto:himmler@eshpm.eur.nl) (SH) and [stockel@eshpm.eur.nl](mailto:stockel@eshpm.eur.nl) (JS)

## Abstract

Decisions on interventions or policy alternatives affecting health can be informed by economic evaluations, like cost-benefit or cost-utility analyses. In this context, there is a need for valid estimates of the monetary equivalent value of health (gains), which are often expressed in € per quality-adjusted life years (QALYs). Obtaining such estimates remains methodologically challenging, with a recent addition to the health economists' toolbox, which is based on well-being data: The well-being valuation approach. Using general population panel data from Germany, we put this approach to the test by investigating several empirical and conceptual challenges, such as the appropriate functional specification of income utility, the choice of health utility tariffs, or the health state dependence of consumption utility. Depending on specification, the bulk of estimated € per QALY values ranged from €20,000–60,000, with certain specifications leading to more considerable deviations, underlining persistent practical challenges when applying the well-being valuation methodology to health and QALYs. Based on our findings, we formulate recommendations for future research and applications.

## KEYWORDS

health valuation, instrumental variable regression, panel data, piecewise regression, quality-adjusted life years, well-being valuation

## JEL CLASSIFICATION

D61, I18, I31, C33, C36

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2021 The Authors. Health Economics published by John Wiley & Sons Ltd.

## 1 | INTRODUCTION

During the ongoing COVID-19 pandemic, many citizens for the first time directly observe scarcity of goods in the health care sector in terms of testing, ventilation, vaccination capacity, and the prioritisation of services under binding capacity constraints. This scarcity and the broader societal consequences of the pandemic has revealed many difficult trade-offs between health and the economy, and between the needs of different patient groups within the health care sector. While the current attention to such matters is unprecedented, policy makers are confronted with many of these trade-offs also in non-pandemic times. To make informed decisions on policy options, however, requires decision makers to weigh up health and economic consequences, aiming to ensure maximum benefit or minimal harm. Welfare economic tools like cost-benefit-analysis can aid decision makers in this process by providing relevant and clear information to openly address the nature of the trade-offs being made (Chilton et al., 2020; Donaldson & Mitton, 2020; Hendren & Sprung-Keyser, 2020).

Cost-benefit analyses entail measuring and valuing gains and losses (benefits and costs) in monetary units, thereby allowing a holistic perspective on societal trade-offs and identifying which policy option is socially most preferred. In the context of interventions and policies affecting population health (though not necessarily aimed primarily at health), cost-benefit analysis therefore requires to obtain estimates on the monetary equivalent value of health, from here onwards denoted as  $v_Q$  (McIntosh, 2010).

In the narrow health care context,  $v_Q$ , depending on the jurisdiction (Rowen et al., 2017), constitutes an important parameter in health technology assessment. There, value for money considerations are often operationalized using cost-utility analysis, where a new technology's costs are compared to its expected health gain, measured using Quality Adjusted Life Years (QALYs) (Neumann et al., 2016). Equation (1) formulates a generalisation of the corresponding decision rule, with  $\Delta Q$  denoting the health gain (in QALYs) and  $\Delta c_t$  the total costs compared to the alternative treatment:

$$\frac{\Delta c_t}{\Delta Q} < v_Q \quad (1)$$

This cost-effectiveness ratio (ICER) is acceptable, if it lies below  $v_Q$  corresponding to one QALY (Brouwer et al., 2019).<sup>1</sup> While the use and empirical foundation of such threshold values within health care vary across jurisdictions (Cameron et al., 2018; Cleemput et al., 2011), estimating the level of  $v_Q$  corresponding to one QALY, also for the purpose of cost-benefit analysis, is challenging and has been attempted using various methods (see background section). In this endeavour, Huang et al. (2018) were the first to conceptualize and apply the well-being valuation approach for estimating a QALY equivalent  $v_Q$ , providing estimates of A\$42,000 (€28,000) to A\$67,000 (€45,000). This method is based on the marginal rate of substitution between income and health. Further exploration of the approach is needed to be able to judge whether the corresponding estimates are indeed helpful for informing  $v_Q$ . This paper aims to make the following contributions: Firstly, by applying a similar approach as Huang et al. (2018) and using data from a different context, we generate further insights regarding the validity and reliability of the well-being valuation method for determining  $v_Q$ . Secondly, we aim to address some empirical and methodological challenges associated with applying the well-being valuation method in general and for valuing QALYs in particular, which were not fully addressed in previous studies. By using German data, an additional contribution lies in providing information on  $v_Q$  for a context in which such estimates are scarce, a result of German health authorities not (explicitly) basing their reimbursement decisions on the framework outlined in Equation (1).<sup>2</sup>

We used data from the Socio-Economic Panel (2019), or SOEP, from 2002 to 2018. Fixed-effects and instrumental variable regressions were used to address endogeneity concerns regarding the impact of income on life satisfaction. Our baseline estimates indicate population average monetary valuations of a QALY of €22,717 and €58,533, with and without instrumenting for income. However, alternative specifications and robustness checks lead to varying estimates, highlighting the empirical challenges and the consequences of methodological choices on the obtained monetary values, and areas for future research.

## 2 | THE SEARCH FOR $v_Q$ AND THE WELL-BEING VALUATION METHOD

Various methods have been used in the ongoing endeavour of obtaining estimates of  $v_Q$ , producing a range of conceptually different values. One approach, employed by Mason et al. (2009), bases  $v_Q$  on estimates of the value

of preventing a statistical fatality, a concept commonly used in public sector safety policies. Another approach calculating  $v_Q$  entails using relative risk aversion in relation to income (Phelps, 2019). However,  $v_Q$  estimates have predominantly been obtained based on stated preferences, by asking individuals directly about their willingness to pay (WTP) for specific health gains. Ryen and Svensson (2015) summarized the extensive literature that used WTP methods to identify  $v_Q$  and reported trimmed mean and median estimates of €74,159 and €24,226 (in 2010 price levels).

Huang et al. (2018) proposed an alternative method for estimating  $v_Q$ , based on revealed, although subjective, information: the well-being valuation approach. This method has been applied to obtain monetary valuations for various other non-market goods, including specific health outcomes and diseases (Brown, 2015; Ferrer-i-Carbonell & van Praag, 2002; Howley, 2017; McNamee & Mendolia, 2018), informal care provision (McDonald & Powdthavee, 2018; van den Berg & Ferrer-i-Carbonell, 2007), air pollution and natural disasters (Luechinger, 2009; Luechinger & Raschky, 2009), national security (Frey et al., 2009) or the welfare effects of sports events (Dolan et al., 2019). In their study, Huang et al. (2018) used data from the HILDA panel survey from Australia and obtained  $v_Q$  estimates of A\$42,000 (€28,000) to A\$67,000 (€45,000), which were similar to threshold values applied for funding decisions in Australia. Recently, Himmler et al. (2020) applied the well-being valuation approach in a cross-sectional sample from the UK to estimate  $v_Q$ , as well as an equivalent value for broader well-being. They report a base case  $v_Q$  estimate of £30,786 (approximately €35,000).

Both stated preference WTP and well-being valuation approaches have advantages and disadvantages and may answer different questions based on how  $v_Q$  is specified. The former allows researchers to tailor their experimental design to specific contexts and control for undesired influences. For instance, WTP can be expressed from an individual or societal perspective (Bobinac et al., 2013), capturing more than self-interested motivations when establishing WTP-based  $v_Q$  estimates. Similarly, equity concerns relating to specific health states or streams (Dolan & Olsen, 2001; Pinto-Prades et al., 2014), but also socio-economic health inequalities can be connected with the QALY framework (Wagstaff, 1991). Furthermore, one can also pose WTP questions from an *ex-ante* or *ex-post* perspective, with the former having the advantage of capturing options value (Gyrd-Hansen, 2003; Philipson & Jena, 2006). However, the practice of asking individuals directly for the value of a prospect brings unique challenges; hypothetical response bias and insensitivity to scope or framing effects are only some of the practical concerns (see Kling et al. (2012)) that have been found to apply when obtaining WTP estimates for a QALY (Ahlert et al., 2016; Bobinac et al., 2012; Gyrd-Hansen et al., 2014; Soeteman et al., 2017).

The well-being valuation approach avoids these challenges by relying on (usually) large-scale observational data, promising to provide a more inclusive picture of the range of preferences over health and wealth across diverse sub-populations. However, the approach limits the scope to respondents' individual *ex-post* valuations, while endogeneity concerns are a prevailing issue as it relies on the estimation of causal effects of health and income to calculate trade-offs.

### 3 | METHODS

#### 3.1 | Conceptual framework

We generally followed the framework proposed by Huang et al. (2018) for obtaining  $v_Q$ . In a simplified model, the subjective well-being (SWB) of individual  $i$  at time  $t$ , as a proxy for individual utility, is assumed to be described by:

$$W_{it} = W(Y_{it}, H_{it}) \quad (2)$$

where  $W_{it}$  is a vector of the individual's well-being at all observed time points ( $w_{it}$ ),  $Y_{it}$  is the corresponding incomes ( $y_{it}$ ), and  $H_{it}$  a vector of health states ( $h_{it}$ ). The total well-being experienced by individual  $i$  over a time interval of length  $T$  can then be described by a simple cumulative sum of individual well-being states across time;

$$W_i = \sum_{t=0}^T W(Y_{it}, H_{it}) \quad (3)$$

Within this framework, consider an individual experiencing a change to their health vector  $\Delta H_i$  within the time window  $T$ . For the individual to remain on the same level of SWB  $W_i$  requires an offsetting income change  $\Delta Y_i$ ;

$$W_i = W(Y_i + \Delta Y_i, H_i + \Delta H_i) \quad (4)$$

The proposed approach estimates the population average  $\Delta Y$  necessary to offset an imposed hypothetical health state change  $\Delta H$  over  $T$  equivalent to one QALY. Therefore,  $\Delta Y$  is the compensating income variation (CIV) for one QALY, or short  $CIV_{QALY}$ .

### 3.2 | Baseline specification

Following Huang et al. (2018), an ordinary least squares (OLS) fixed-effects regression was estimated to calculate the impact of health and income on SWB within a time window  $T$  of 2 years ( $t_0$  and  $t_{-1}$ ). Modelling SWB as linear despite the cardinal nature of life satisfaction is a widely used approach, see for example, Ferrer-i-Carbonell and van Praag (2002). The underlying empirical model takes the following form;

$$W_{irt} = \alpha + \beta_0 H_{irt} + \beta_1 H_{irt-1} + \delta_0 Y_{irt} + \delta_1 Y_{irt-1} + \tau X_{irt} + \lambda_i + \mu_r + \epsilon_t + u_{irt} \quad (5)$$

where  $W_{irt}$  refers to the SWB of individual  $i$  living in region  $r$  at time  $t$ , measured using life satisfaction data. The individual's health status  $H_{irt}$  is captured by health utility values based on the short form six dimensions (SF-6D) instrument and its UK utility tariff (Brazier & Roberts, 2004). Household income is denoted by  $Y_{irt}$ . Lagged variables of health and income were included to not be limited to short-term one-year changes and to partly account for reverse causality. We control for a vector  $X_{irt}$  of other potential time-varying confounders. To account for time-invariant unobservables, we incorporated individual ( $\lambda_i$ ), state ( $\mu_r$ ), and time ( $\epsilon_t$ ) fixed-effects.  $u_{irt}$  denotes the error term. Heteroscedasticity-robust standard errors were used in all estimations.

In a second step, we obtained  $CIV_{QALY}$  values by dividing the health status coefficients ( $\beta_0$  and  $\beta_1$ ) by the income coefficients ( $\delta_0$  and  $\delta_1$ ):

$$CIV_{QALY} = \frac{\beta_0 + \beta_1}{\delta_0 + \delta_1} \quad (6)$$

The corresponding values represent the marginal rate of substitution between income and health with respect to well-being, based on the overall population average.  $CIV_{QALY}$  thereby is the empirical conceptualisation of  $v_Q$  using the well-being valuation approach. Income outliers (as will be defined in Section 3.4.1) were dropped from the baseline analysis.

### 3.3 | Instrumental variable specification

A well-documented problem of the well-being valuation approach is the endogeneity of the income coefficient estimate. This was frequently addressed using an instrumental variable (IV) (see e.g., Brown, 2015; Howley, 2017; McNamee & Mendolia, 2018). Huang et al. (2018) instrumented income with the occurrence of financial-worsening-events such as personal bankruptcy or large financial losses.

Lacking such information, we followed Luechinger (2009), who used predicted labour-market earnings based on industry-occupation cells as income instrument. The rationale is that shifts in predicted income correspond to industry and/or occupation wide trends, which correlate with the development of negotiated wages or collective wage agreements, but do not reflect individual-level effort or circumstances. Further, it is assumed that the income variance across industries and occupations captures information on the unobserved costs of income generation such as stress and/or associated health risks, and that unobserved selection effects of certain types of individuals into industries and

occupations are captured in the time-invariant fixed-effects. One advantage of this instrument is that the captured income shifts have a rather permanent nature, whereas financial-worsening-events or lottery wins can be highly transitory shocks. In addition permanent income shifts have been found to be of higher relevance for individuals' well-being (Bayer & Juessen, 2015; Cai & Park, 2016).

The identifying assumption is, therefore, that income variation across industries and occupations over time is uncorrelated with individual-level characteristics and especially life satisfaction, besides the effect of income changes themselves. To implement the IV approach we followed a two-stage least squares estimation procedure. In a first step we estimated the individual's labour market earnings  $L_{irt}$  based on the following regression;

$$L_{irt} = \alpha + \rho_0 I_{irt} + \rho_1 O_{irt} + \rho_2 T_{irt} + \rho_3 R_{irt} + \mu_r + \epsilon_t + u_{irt} \quad (7)$$

from which we obtained fitted values, constituting the predicted labour earning conditional on the individual's industry-occupation cell ( $I_{irt}$  and  $O_{irt}$ ), work tenure ( $T_{irt}$ ), and work-hours ( $R_{irt}$ ) and a set of industry- and year-fixed-effects.

The obtained predicted labour earnings were summed on the household level and weighted by household composition to obtain the predicted household labour income  $\hat{L}_{irt}^{HH}$ , the instrument used in the first-stage regression;

$$Y_{irt} = \bar{\alpha} + \bar{\beta}_0 H_{irt} + \bar{\beta}_1 H_{irt-1} + \bar{\delta}_0 \hat{L}_{irt}^{HH} + \bar{\delta}_1 \hat{L}_{irt-1}^{HH} + \bar{\tau} X_{irt} + \bar{\lambda}_i + \bar{\mu}_r + \bar{\epsilon}_t + \bar{u}_{irt} \quad (8)$$

from which we obtained the fitted values for individual income,  $\hat{Y}_{irt}$ . In the second stage we substituted income  $Y_{irt}$  by  $\hat{Y}_{irt}$ , estimating

$$W_{irt} = \alpha^I + \beta_0^I H_{irt} + \beta_1^I H_{irt-1} + \delta_0^I \hat{Y}_{irt} + \delta_1^I \hat{Y}_{irt-1} + \tau^I X_{irt} + \lambda_i^I + \mu_r^I + \epsilon_t^I + u_{irt}^I. \quad (9)$$

The resulting coefficients for health ( $\beta_0^I$  and  $\beta_1^I$ ) and income ( $\delta_0^I$  and  $\delta_1^I$ ) were then included in Equation (6) to calculate the IV  $CIV_{QALY}$  estimate. For further details please see Supplementary Appendix A3.

### 3.4 | Alternative model specifications

#### 3.4.1 | Treatment of outliers

Due to a right-skewed and long-tailed income distribution, with self-reported income often misreported or even exaggerated (Hariri & Lassen, 2017), income outliers may have a large effect on  $CIV_{QALY}$  estimates when using linear models (Rousseeuw & Leroy, 1987). To identify outliers, which remains challenging for fixed-effects models (Verardi & Croux, 2009), we reformulated our base case model as a pooled OLS model and calculated DFbeta, a measure quantifying the impact that dropping an observation has on the coefficient estimate. All observations with a DFbeta larger than 1, the recommended threshold (Bollen & Jackman, 1985), were dropped from the baseline analysis. In a robustness check we repeated the calculations including these outliers.

#### 3.4.2 | Income specification

To accommodate the diminishing marginal return of income we log-transformed income (Layard et al., 2008).  $CIV_{QALY}$  was then estimated based on a slightly modified equation as used by Ólafsdóttir et al. (2020) and van den Berg and Ferrer-i-Carbonell (2007). This entailed dropping the lagged income and health coefficients as used in our base model (Equation 6).

$$CIV_{QALY} = \bar{y} * \left( \exp\left(\frac{-\beta_0 * \frac{1}{\Delta}}{\delta_0}\right) - 1 \right) * \Delta \quad (10)$$

In the log-income specification  $CIV_{QALY}$  was calculated as the percentage share of annual income (median annual income  $\bar{y}$ ). By construction,  $CIV_{QALY}$  values would be confined to be no greater than this income level which may be acceptable when valuing small gains or changes but not a full QALY. Therefore, we added the parameter  $\Delta$  to the equation and set it to 10. Instead of calculating the monetary equivalent of a one QALY change we calculated the equivalent of a 0.1 QALY change and multiplied it by 10.

To account for the non-linearity of income without imposing a logarithmic functional form, which may not adequately capture the relationship especially on the lower end of the income distribution, we furthermore tested a piecewise linear specification similar to Ólafsdóttir et al. (2020). To obtain the appropriate number of income splines and cut-off values, we iteratively combined income-deciles. The equality of coefficient estimates of adjacent splines was tested and non-significantly different splines were gradually combined until coefficients were significantly different and model fit did not improve.  $CIV_{QALY}$  values were then calculated for each income spline and also aggregated by weighting according to the number of individuals in the respective splines. Estimating a piecewise IV specification was not feasible, as one distinct income instrument would have been required for each of the splines.

### 3.4.3 | Choice of utility tariff

Lacking a German specific SF-6D utility tariff we relied on the UK time-trade-off based value set (Brazier & Roberts, 2004) to construct health utilities. In an alternative specification we explored the importance of tariff choice by instead applying a recently developed value set from the Netherlands which was estimated using a discrete choice experiment (Jonker et al., 2018).

### 3.4.4 | Health state dependence of the utility of consumption

Another empirical issue of concern relates to the interaction between health and income and experienced (consumption) utility. This so-called health state dependence implies that the marginal utility gain from a given income change is directly dependent on the underlying health status (Finkelstein et al., 2013). So far, there is only inconclusive evidence on the magnitude and the direction of this effect: Finkelstein et al. (2013) found a negative health state dependence, a higher marginal utility of income in good compared to bad health, based on US data. However, replicating their approach using European data, Kools and Knoef (2019) found evidence for positive health state dependence, potentially due to differing provision of public goods in European healthcare systems.

As illustrated by both Finkelstein et al. (2013) and Kools and Knoef (2019), health state dependence has important implications for (health) economic issues such as the optimal design of insurance contracts or individual-level decisions on life-cycle savings. In the context of estimating  $CIV_{QALY}$ , which requires a simultaneous measurement of the well-being impacts of both health and income separately, a thorough investigation of the life-cycle development of health states and the associated changes in consumption utility seems warranted.

To explore the potential impact of health state dependence on  $CIV_{QALY}$  estimates, we reduced our sample to those individuals that transitioned between health states. Finkelstein et al. (2013) used the onset of chronic diseases for this purpose. While this represents a convenient definition for an elderly population, we took a different approach, allowing us to observe the transition of individuals from good to bad health also for healthier groups. First, we reduced the sample to individuals whose mental or physical short form health questionnaire (SF-12) component scores changed by at least 10, or one standard deviation, throughout their respective observation period.<sup>3</sup> This was done to ensure that individuals in this group have experienced a consequential change in their mental and/or physical health. Good health states were defined as periods in which either of the two scores was above their respective individual-level mean; bad health states if they were below. Secondly, we conditioned on the consecutive observation of differing health states with at least two consecutive periods needed to be observed in either state. This allowed us to estimate  $CIV_{QALY}$  for good and bad health separately while also ensuring that individuals transition into longer-term health states (see Supplementary Appendix A4 for details). Importantly, the sample included individuals transitioning from good to bad health and vice versa, although the former is most frequent.

## 4 | DATA

We used data from the annual SOEP panel survey, providing a representative sample of the adult (aged 16+) German population (Goebel et al., 2019). Ethical approval with respect to the surveying process generating the underlying data was obtained by the SOEP researchers directly. SF-6D health utilities were constructed from SF-12 data, which is biennially included in the survey since 2002. To facilitate the specified two-year time-frame  $T$  used for the  $CIV_{QUALY}$  calculations, and to prevent dropping observations from every second year, we linearly imputed SF-6D values for intermediate years. However, this was only done if individuals were observed for three consecutive years with two completed SF-12 surveys.

Life satisfaction was measured on a 10-point scale ranging from 0 (“completely dissatisfied”) to 10 (“completely satisfied”). Information on individuals' income was based on self-reported monthly net household income. To account for differences in household composition, we calculated equivalized household income, following the definition by Hagenaaers et al. (1994). Income data was converted to 2018 prices using the official consumer price indices (Federal Statistical Office, 2020).

To construct our instrument we extracted information on net labour income and individuals' industry and occupation. We dropped households with individuals where information on labour income but not on industry/occupation was available. Predicted labour income was assumed to be zero for all individuals with no labour income information, or who stated that they were not employed.<sup>4</sup>

We furthermore extracted information on a similar set of variables as used by Huang et al. (2018) to control for confounding factors. These included age, disability, marital status, employment status, educational attainment and leisure time. Table 1 summary statistics of the analysis data, consisting of 29,735 individuals providing 186,906 individual-year observations. Supplementary Table A1.1 provides an overview of the conditioning applied to the SOEP data, while Supplementary Table A1.2 shows that the sub-sample of employed individuals who were dropped because of missing industry/occupation information is comparable to the remaining sample of employed individuals. As the exclusion of individuals without at least two consecutive SF-6D values was the only major selection criterion, the sample remained largely representative for the overall German population.

TABLE 1 Descriptive statistics

Variable	Mean	Std. Dev.	Description
Life satisfaction	7.09	1.71	0 (lowest) to 10 (highest)
Income in 1000's	2.03	1.29	Monthly household income in €
SF-6D utility	0.73	0.13	0.345-1, 1 perfect health
Disability	0.14	0.35	1 if disability status
Age in years	53.67	15.78	
(de facto) Married	0.67	0.47	1 if married, living together
Education: Primary	0.12	0.32	1 if primary educated
Education: Tertiary	0.63	0.48	1 if secondary educated
Education: Secondary	0.25	0.43	1 if tertiary educated
Leisure time	2.18	2.03	Hours per day
Employed	0.56	0.50	1 if employed
Unemployed	0.04	0.21	1 if unemployed
Work hours	21.22	20.99	Hours per week
Tenure	7.03	9.96	Years at current job
Individuals * years		186,902	
Individuals		29,735	

Source: Own calculations based on SOEP Waves 2002–2018.

## 5 | RESULTS

## 5.1 | Baseline results

The baseline OLS and IV results, are shown in Table 2, separating between results using the full dataset with imputed SF-6D values, and the dataset without imputation. To construct our instrumental variables, we predicted labour incomes based on industry/occupation for 125,229 observations. Supplementary Appendix A3 provides details on this prediction and the associated errors, which were small for the largest part of the income distribution. The instruments were significant in the first stage regression (Supplementary Table A3.1) and passed the Cragg-Donald weak identification test ( $F$ -value: 1864 and 192). This indicates a high relevance of the instrument, a common finding for this type of instrument (Bayer & Juessen, 2015; Luechinger, 2009). The Hausman test for endogeneity of the instrumented variables was significant, signalling that income should not be treated as exogenous.

Equivalized monthly household income, health status (SF-6D utility), and their lagged values were positive and significant predictors of life satisfaction in the OLS specification. This was also the case when instrumenting for income,

TABLE 2 Baseline results

	SF-6D imputation				No imputation			
	OLS		IV		OLS		IV	
Income in 1000's	0.05***	(0.01)	0.10***	(0.03)	0.05***	(0.01)	0.14***	(0.05)
Income in 1000's ( $t - 1$ )	0.01	(0.01)	0.04	(0.03)	-0.00	(0.01)	-0.00	(0.07)
SF-6D utility	3.12***	(0.06)	3.12***	(0.05)	3.52***	(0.06)	3.51***	(0.05)
SF-6D utility ( $t - 1$ )	0.10*	(0.06)	0.10*	(0.05)	0.47***	(0.05)	0.46***	(0.05)
Disability	-0.14***	(0.02)	-0.14***	(0.02)	-0.09***	(0.03)	-0.09***	(0.02)
Age	0.09***	(0.01)	0.08***	(0.02)	0.05***	(0.01)	0.05***	(0.01)
Age squared	-0.00***	(0.00)	-0.00**	(0.00)	-0.00**	(0.00)	-0.00	(0.00)
(de facto) Married	0.18***	(0.02)	0.18***	(0.02)	0.17***	(0.03)	0.16***	(0.02)
Primary education	-0.18*	(0.09)	-0.21***	(0.08)	-0.10	(0.15)	-0.13	(0.13)
Tertiary education	-0.18***	(0.06)	-0.19***	(0.05)	-0.19***	(0.07)	-0.20***	(0.07)
Leisure time	0.03***	(0.01)	0.03***	(0.00)	0.03***	(0.01)	0.03***	(0.01)
Leisure time squared	-0.00***	(0.00)	-0.00***	(0.00)	-0.00**	(0.00)	-0.00***	(0.00)
Unemployed	-0.52***	(0.03)	-0.53***	(0.02)	-0.53***	(0.04)	-0.53***	(0.03)
Work hours	0.00***	(0.00)	0.00***	(0.00)	0.00***	(0.00)	0.00	(0.00)
Tenure	-0.01***	(0.00)	-0.01***	(0.00)	-0.01***	(0.00)	-0.01***	(0.00)
Model statistics								
Cragg-Donald			1864				192	
Anderson			3642				382	
Endogeneity test			10.0				5.8	
BIC	540,754.8		540,994.6		250,099.1		236,537.9	
Observations	186,902		186,902		93,450		85,433	
Individuals	29,735		29,735		29,735		21,718	
CIV-QALY in €	58,533		22,717		80,522		28,130	

Source: Own calculations based on SOEP Waves 2002–2018.

Abbreviations: BIC, Bayesian information criteria; CIV-QALY, compensating income variation of one QALY; IV, instrumental variable; OLS, ordinary least squares.

\* $p < 0.10$ , \*\* $p < 0.05$ , \*\*\* $p < 0.01$ .



except that the lagged income coefficient was insignificant. We observed a two-fold increase in the income coefficients in the IV model (0.048 vs. 0.098), a similar magnitude to what has been observed in previous studies using the SOEP (Bayer & Juessen, 2015; Pischke, 2011). Interestingly, the difference is minimal compared to what was observed by Huang et al. (2018), who reported an IV coefficient which was 130 times larger than the OLS coefficient (0.080 and 0.0006). Applying the estimated income and SF-6D coefficients to Equation (6) resulted in a  $CIV_{QALY}$  value of €58,533 in the OLS model and €22,717 when instrumenting for income. This value represents the average amount of additional income necessary to maintain the same level of life satisfaction if a hypothetical health change of one QALY is imposed.

Without SF-6D imputation, reducing our sample to 85,433 observations across 21,718 individuals, the OLS results increased by a factor of 1.38 to €80,522 while the IV-based value increased by a factor of 1.24 to €28,130. These differences were driven by larger SF-6D and income coefficients compared to the baseline calculations, possibly resulting from increased within-person variance as the distance between observations is two years instead of one. For the remainder of the results presented, we will be using the full dataset with imputed SF-6D values to make use of the largest amount of information available.

Table 3 columns 2–3 contains estimates for East and West Germany separately, motivated by the persisting differences in life satisfaction and income levels (Frijters et al., 2004; Vatter, 2020). OLS-based  $CIV_{QALY}$  estimates were €75,748 in the West and €28,548 in the East. The IV-based estimate was also higher in the West compared to the East (€20,750 and €12,982), although the relative difference was lower (factor of 3.64 and 2.20). In both models, this difference was mainly driven by a considerably larger income coefficients in the East, likely due to the prevailing income differences between West and East; observed average monthly equalized income was €2140 in the West and only €1652 in the East.

We investigated the (undesired) impact of macro economic conditions on  $CIV_{QALY}$  estimates by excluding the years of the financial crisis and recession in Germany (2007–2009). As shown in Table 4 (columns 4–6), this had only a minor impact on the OLS and IV  $CIV_{QALY}$  values (€54,567 and €20,574). However, estimates based on the pre-crisis time periods 2002–2006 (€56,640 and €7,720) were substantially lower compared to estimates based on data from 2010–2018 (€70,572 and €24,811). This resulted from larger estimated effects of income in earlier periods, which may both be a

TABLE 3 Results by region and time-period

	Baseline		East		West		w/o 2007–2009		2002–2006		2010–2018	
	OLS	IV	OLS	IV	OLS	IV	OLS	IV	OLS	IV	OLS	IV
Income in 1000's	0.05*** (0.01)	0.10*** (0.03)	0.13*** (0.02)	0.18** (0.08)	0.04*** (0.01)	0.07** (0.04)	0.05*** (0.01)	0.11*** (0.04)	0.06*** (0.01)	0.29*** (0.09)	0.04*** (0.01)	0.09* (0.05)
Income in 1000's ( $t - 1$ )	0.01 (0.01)	0.04 (0.03)	0.00 (0.02)	0.03 (0.06)	0.01 (0.01)	0.04 (0.03)	0.01* (0.01)	0.05 (0.03)	−0.00 (0.01)	0.10 (0.08)	0.01 (0.01)	0.04 (0.04)
SF-6D utility	3.12*** (0.06)	3.12*** (0.05)	2.90*** (0.13)	2.90*** (0.12)	3.18*** (0.07)	3.17*** (0.07)	3.16*** (0.07)	3.15*** (0.07)	2.93*** (0.15)	2.92*** (0.15)	3.08*** (0.08)	3.08*** (0.08)
SF-6D utility ( $t - 1$ )	0.10* (0.06)	0.10* (0.05)	−0.12 (0.12)	−0.12 (0.12)	0.16** (0.07)	0.16** (0.07)	0.10 (0.07)	0.09 (0.07)	0.06 (0.14)	0.06 (0.14)	−0.07 (0.08)	−0.07 (0.08)
Model statistics												
Cragg-Donald	1863.7		323.9		680.2		783.4		181.2		494.3	
Anderson	3642.0		544.4		1265.5		1429.5		328.8		907.3	
Endogeneity test	10.0		1.5		5.8		9.7		8.2		2.7	
BIC	540,755	540,995	127,072	127,092	412,723	412,877	431,238	431,487	129,869	130,432	276,374	276,464
Observations	186,902	186,902	43,447	43,447	143,361	143,361	151,461	151,461	48,678	48,678	101,048	101,048
CIV-QALY in €	58,533	22,717	20,750	12,982	75,748	28,548	54,567	20,574	56,640	7,720	70,572	24,811

Source: Own calculations based on SOEP Waves 2002–2018.

Abbreviation: BIC, Bayesian information criteria; CIV-QALY, compensating income variation of one QALY; IV, instrumental variable; OLS, ordinary least squares.

$p < 0.10$ , \*\* $p < 0.05$ , \*\*\* $p < 0.01$ .

TABLE 4 Income specifications

	Baseline		With Outliers		Log income		Piecewise OLS
	OLS	IV	OLS	IV	OLS	IV	
Income in 1000's	0.05*** (0.01)	0.10*** (0.03)	0.03*** (0.01)	0.10*** (0.03)			
Income in 1000's ( $t - 1$ )	0.01 (0.01)	0.04 (0.03)	0.01*** (0.00)	0.04 (0.03)			
SF-6D utility	3.12*** (0.06)	3.12*** (0.05)	3.12*** (0.06)	3.12*** (0.06)	3.18*** (0.05)	3.16*** (0.05)	3.18*** (0.05)
SF-6D utility ( $t - 1$ )	0.10* (0.06)	0.10* (0.05)	0.10* (0.06)	0.10* (0.06)			
Log income					0.24*** (0.02)	0.63*** (0.13)	
1st income spline							0.43*** (0.05)
2nd income spline							0.27*** (0.05)
3rd income spline							0.11*** (0.02)
4th income spline							0.01 (0.01)
Model statistics							
Cragg-Donald		1863.7		825.8		1329.9	
Anderson		3642.0		1529.4		1278.2	
Endogeneity test		10.0		12.9		9.7	
BIC	540,755	540,995	540,801	541,306	540,506	541,501	540,448
Observations	186,902	186,902	186,906	186,906	186,902	186,902	186,902
CIV-QALY in €	58,533	22,717	82,484	22,782	153,877	81,649	97,486
Without 4th spline							19,515

Note: Instrumental variable did not pass weak identification tests for piecewise income specification. CIVs for piecewise regression represents population-weighted averages of all splines or the first three splines (€7,347, €11,686, €29,548 and €409,810).

Source: Own calculations based on SOEP Waves 2002–2018.

Abbreviations: BIC, Bayesian information criteria; CIV-QALY, compensating income variation of one QALY; IV, instrumental variable; OLS, ordinary least squares.

\* $p < 0.10$ , \*\* $p < 0.05$ , \*\*\* $p < 0.01$ .

result of a positive trend in incomes or a shift in population preferences and values over the last decades. Supplementary Table A2.1 provides further results on age and gender subgroups.

## 5.2 | Specifications related to income

Re-estimating our baseline models including four individual-year observations which were flagged as outliers lead to a considerably lower income coefficient in the OLS model (Table 4 columns 3–4). This increased the  $CIV_{QALY}$  value to €82,484. The IV estimates were only minimally affected by this (€22,782). The outlier observations corresponded to two

individuals from the same household, which reported a drop in monthly income from €142,534 to €14,051 within 2 consecutive years, while reporting constant life satisfaction.

In the models using log-transformed income (Table 4 columns 5–6), the income coefficient was 0.24, larger than reported before by Pischke (2011) (0.125 to 0.182), who also used the SOEP. The corresponding IV coefficient, with a value of 0.63, was on the higher end of previous IV estimates based on the industry-wage structure and the SOEP: Luechinger (2009) reported an estimate of 0.55, while Pischke (2011) reported values ranging from 0.489 to 0.617. Previous estimates based on instruments using lagged or future income shocks were also similar, with Bayer and Juessen (2015) providing a range of 0.45 to 0.50 for permanent income shifts.<sup>5</sup> The log-transformation resulted in considerably larger  $CIV_{QALY}$  values compared to the baseline. The OLS values increased by a factor of 2.63 to €153,877 while the IV values increase by a factor of 3.59 to €81,649.<sup>6</sup>

The piecewise linear specification was estimated with ultimately four income splines. The cut-off points were at the 20<sup>th</sup> percentile (€1200), the 40<sup>th</sup> percentile (€1546), and the 80<sup>th</sup> percentile (€2635). Figure 1 plots the overall distribution of life satisfaction across income, and the linear fit of life satisfaction across splines, indicating a non-linear, diminishing pattern. The spline-specific  $CIV_{QALY}$  values were €7347, €11,686, €29,548 and €409,810. The population aggregated  $CIV_{QALY}$  was €97,486. This estimate was driven by the large  $CIV_{QALY}$  value in the fourth income spline, where the income coefficient was insignificant. Using the three significant splines lead to a  $CIV_{QALY}$  value of €19,515.

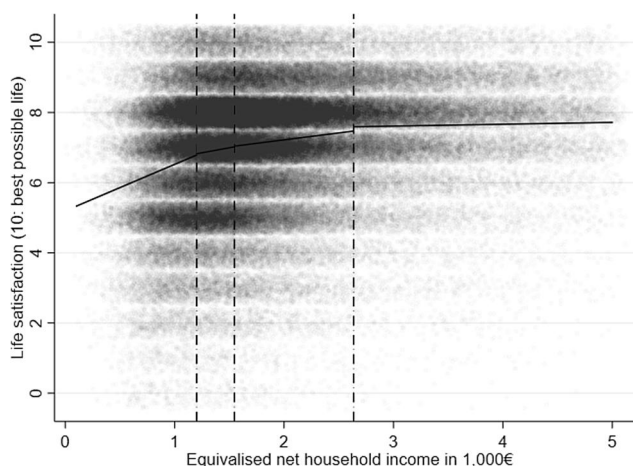
### 5.3 | Specifications and issues related to health

Choice of SF-6D value set.

Applying the Dutch SF-6D value set shifted the distribution of health utilities (Figure 2), with the mean utility decreasing from 0.725 to 0.554. These differences likely reflect methodological differences rather than actual variation in health state preferences between the UK and the Netherlands as UK and Dutch tariffs for the EQ-5D have been shown to be similar (Norman et al., 2009).

The estimated  $CIV_{QALY}$  values using the Dutch SF-6D tariff were markedly smaller (Table 5). The OLS estimates decreased from €58,533 to €32,534, while the IV estimates decreased from €22,717 to €13,054. This shift was caused by the smaller SF-6D coefficients (3.12 to 1.78), resulting from the wider spread of the Dutch tariff, which ranges from –0.44 to 1, allowing for negative health state utility, instead of 0.345 to 1 as in the UK value set. The same actual change in health corresponds to a larger change in SF-6D utility in the Dutch tariff which reduces the impact of a (hypothetical) one unit change in SF-6D on life satisfaction.

Health state dependence of the utility of consumption



**FIGURE 1** Relationship between life satisfaction and income across income splines. Life satisfaction values are depicted as small grey dots. Black dash-dotted vertical lines represent the income splines used in the piece-wise linear regression. Black horizontal lines plot the linear fit within these splines

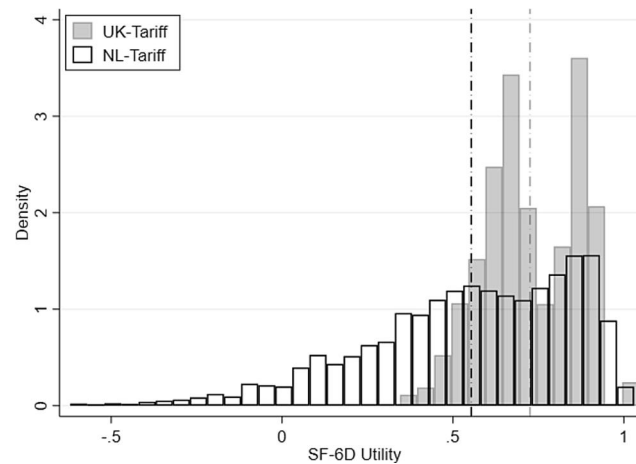


FIGURE 2 SF12 index values using UK and Dutch tariffs. The black dash-dotted line indicates the Dutch tariff mean. The grey dash-dotted line indicates the UK tariff mean. The distributions and means reflect SF-6D values based on self-reported SF12 questionnaires only

	UK Tariff		Dutch Tariff	
	OLS	IV	OLS	IV
Income in 1000's	0.05*** (0.01)	0.10*** (0.03)	0.05*** (0.01)	0.09*** (0.03)
Income in 1000's ( $t - 1$ )	0.01 (0.01)	0.04 (0.03)	0.01 (0.01)	0.05* (0.03)
SF-6D utility	3.12*** (0.06)	3.12*** (0.05)	1.78*** (0.03)	1.78*** (0.03)
SF-6D utility ( $t - 1$ )	0.10* (0.06)	0.10* (0.05)	0.05 (0.03)	0.05 (0.03)
Model statistics				
Cragg-Donald		1863.7		907.1
Anderson		3642.0		1671.4
Endogeneity test		10.0		9.4
BIC	540,755	540,995	538,297	538,523
Observations	186,902	186,902	186,902	186,902
CIV-QALY in €	58,533	22,717	32,534	13,054

TABLE 5 Choice of SF-6D tariffs

Source: Own calculations based on SOEP Waves 2002–2018.

Abbreviations: BIC, Bayesian information criteria; CIV-QALY, compensating income variation of one QALY; IV, instrumental variable; OLS, ordinary least squares.

\* $p < 0.10$ , \*\* $p < 0.05$ , \*\*\* $p < 0.01$ .

We explored the potential impact of health state dependence on  $CIV_{QALY}$  estimates by restricting our sample to individuals experiencing a substantial health change, and splitting their respective observation periods into good and bad health states (see Section 3.4.4). The resulting sample was considerably smaller, including only 5112 individuals yielding 48,861 observations. Nevertheless, the summary statistics suggests that the sample is still comparable to the full population sample (see Supplementary Table A4.1). Table 6 depicts the corresponding estimation results. Compared to the baseline estimates using the full sample,  $CIV_{QALY}$  values based on the combined good and bad health state samples were lower in the OLS model (€39,482) and similar in the IV specification (€20,377). For “good health states”, the corresponding  $CIV_{QALY}$  estimates were lower with €33,336 and €16,532. For “bad health states”, the OLS-based  $CIV_{QALY}$

TABLE 6 Health state dependence

	Baseline		Good health		Bad health	
	OLS	IV	OLS	IV	OLS	IV
Income in 1000's	0.07*** (0.01)	0.17** (0.07)	0.05*** (0.02)	0.11 (0.08)	0.08** (0.04)	0.32 (0.24)
Income in 1000's ( $t - 1$ )	0.03** (0.01)	0.02 (0.06)	0.03** (0.01)	0.05 (0.06)	0.03 (0.03)	0.05 (0.17)
SF-6D utility	3.62*** (0.11)	3.60*** (0.09)	2.51*** (0.14)	2.50*** (0.12)	4.10*** (0.38)	4.03*** (0.37)
SF-6D utility ( $t - 1$ )	0.10 (0.10)	0.11 (0.10)	0.12 (0.12)	0.12 (0.11)	0.32 (0.26)	0.32 (0.27)
Model statistics						
Cragg-Donald		620.7		425.1		95.9
Anderson		1208.4		828.1		188.4
Endogeneity test		3.0		1.8		1.0
BIC	150,481	150,558	102,463	102,497	37,832	37,899
Observations	48,861	48,861	35,401	35,401	13,460	13,460
CIV-QALY in €	39,482	20,377	33,336	16,532	38,374	11,779

Source: Own calculations based on SOEP Waves 2002–2018.

Abbreviations: BIC, Bayesian information criteria; CIV-QALY, compensating income variation of one QALY; IV, instrumental variable; OLS, ordinary least squares.

\* $p < 0.10$ , \*\* $p < 0.05$ , \*\*\* $p < 0.01$ .

estimate was €38,374 and the IV-based estimate €11,779. Important to note is that the drop in the IV based results for the bad health state primarily resulted from a larger income coefficient estimate, even though the SF-6D coefficients increased considerably. These results indicate that there is a positive health state dependence of income in line with the results for Germany by Kools and Knoef (2019). Unfortunately, we were not able to follow Kools and Knoef (2019) and Finkelstein et al. (2013) in focusing on non-working individuals to ensure stable income across health states, ruling out that the increased income coefficients are driven by individuals losing their income, and hence having a larger marginal utility of additional earnings. For our analysis, such a restriction was not feasible, as within-person income variation is necessary to estimate the income coefficients. However, the general empirical pattern remains the same when excluding individuals with large negative income differences between health states (see Supplementary Table A4.2). This also holds when only considering the working population (Supplementary Table A4.3) and those experiencing sudden and severe health changes (Supplementary Table A4.4).

## 5.4 | Robustness checks

Lastly, we tested the robustness of our baseline results to some general concerns regarding our estimation strategy (Table 7). In a first robustness check, we limited our sample to individuals which were in paid employment and provided industry-occupation information, the same sample which was used to obtain estimates for predicted labour income for the IV regression. The resulting OLS-based  $CIV_{QALY}$  was slightly lower than the baseline at €52,829, while the IV-based value was slightly higher than the baseline at €26,097. These differences were driven by the smaller SF-6D coefficients in both OLS and IV models, likely resulting from the working population being healthier as individuals without labour income (the unemployed and retired). The sum of both income coefficients was smaller in the corresponding IV-calculations compared to baseline, increasing the  $CIV_{QALY}$ .

Next, we followed Luechinger (2009) by excluding households with self-employed main income earners, as the income measurement error was likely to be amplified among these individuals. Self-employed individuals are often

TABLE 7 Robustness checks

	Baseline		Working only		No self-employed		No bonus income		With ind/occ missing OLS
	OLS	IV	OLS	IV	OLS	IV	OLS	IV	
Income in 1000's	0.05*** (0.01)	0.10*** (0.03)	0.05*** (0.01)	0.05 (0.03)	0.07*** (0.01)	0.05 (0.04)	0.05*** (0.01)	0.14*** (0.04)	0.04*** (0.01)
Income in 1000's (t - 1)	0.01 (0.01)	0.04 (0.03)	0.01 (0.01)	0.07** (0.03)	0.00 (0.01)	0.08** (0.03)	0.01* (0.01)	0.02 (0.03)	0.01** (0.00)
SF-6D utility	3.12*** (0.06)	3.12*** (0.05)	2.95*** (0.08)	2.94*** (0.07)	2.97*** (0.08)	2.97*** (0.07)	3.12*** (0.07)	3.11*** (0.06)	3.14*** (0.06)
SF-6D utility (t - 1)	0.10* (0.06)	0.10* (0.05)	0.07 (0.07)	0.06 (0.07)	0.01 (0.08)	0.01 (0.07)	0.11* (0.06)	0.11* (0.06)	0.12** (0.06)
Model statistics									
Cragg-Donald		1863.7		1355.7		1898.4		719.1	
Anderson		3642.0		2637.7		3633.0		1334.4	
Endogeneity test		10.0		5.4		7.4		10.1	
BIC	540,755	540,995	319,169	319,323	279,896	280,043	502,827	503,172	578,002
Observations	186,902	186,902	116,125	116,125	101,703	101,703	172,998	172,998	198,950
CIV-QALY in €	58,533	22,717	52,829	26,097	44,058	21,382	53,974	20,464	62,266

Note: Ind/occ refers to specification where individuals without industry/occupation information were included.

Source: Own calculations based on SOEP Waves 2002–2018.

Abbreviations: BIC, Bayesian information criteria; CIV-QALY, compensating income variation of one QALY; IV, instrumental variable; OLS, ordinary least squares.

\* $p < 0.10$ , \*\* $p < 0.05$ , \*\*\* $p < 0.01$ .

reluctant to disclose their income, while also experiencing unstable income streams and hence, even if not reluctant to report, they might simply misreport accidentally. The resulting  $CIV_{QALY}$  estimates and income and SF-6D coefficients were similar to the baseline estimates (€55,359 and €20,352).

Another concern relating to the instrument is that observed income changes may also relate to individual effort, which likely impacts income differently across industries and occupations. Unfortunately, effort cannot be observed. To nevertheless explore this, we use information on reported bonuses, gratifications, or profit sharing to identify the group of individuals for whom this might be a relevant concern, as for them effort would have the highest impact on income and life satisfaction. To test the robustness of our results to this potential bias, we estimate our baseline models excluding such observations. The results in Table 7 columns 7–8 suggest that this bias is relatively limited.

To investigate the potential impact of dropping employed individuals without industry/occupation information (as required for constructing the IV), we included those observations in a further robustness check (Table 7 column 9). The corresponding OLS estimates for income coefficients and  $CIV_{QALY}$  (€62,266) are comparable to our baseline estimates. However, by construction, we cannot confirm this for the IV estimates.

## 6 | DISCUSSION

Applying the well-being valuation approach to longitudinal health and income data from Germany, we estimated the monetary equivalent value of one year in full health  $v_Q$  (equivalent to one QALY). Beyond demonstrating the feasibility of this approach in a new country context, we explored additional empirical and methodological challenges with implications for the practical usefulness of well-being valuation based  $v_Q$  estimates (denoted as  $CIV_{QALY}$ ).

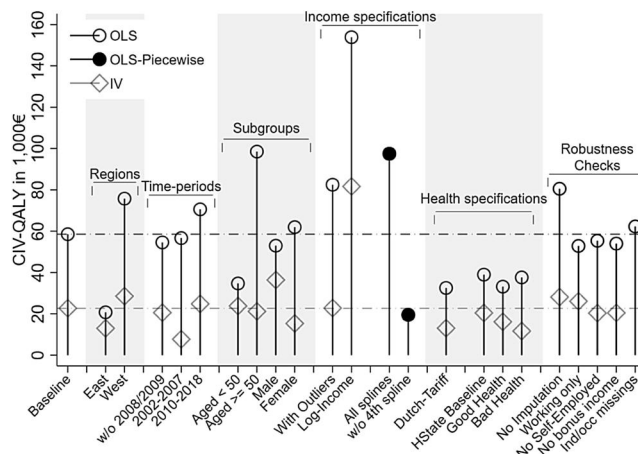


FIGURE 3 Overview of  $CIV_{QALY}$  estimates. The horizontal dash-dotted lines indicate our baseline  $CIV_{QALY}$  estimates from the baseline OLS (black) and IV (grey) specifications

## 6.1 | Overview and context of results

Figure 3 presents an overview of our  $CIV_{QALY}$  estimates. The baseline calculations provided average monetary valuations of a QALY of €58,533 (OLS) and €22,717 (IV).  $CIV_{QALY}$  estimates varied across model specifications with the bulk of values lying between €20,000 and €60,000 and the (OLS) log-income specifications reaching the maximum value of €153,877. Instrumenting for income consistently lead to lower values, a common finding in the well-being valuation literature (e.g., Ólafsdóttir et al. (2020)).

The range of  $CIV_{QALY}$  estimates obtained in our study fit into the ballpark of more reasonable stated preference estimates (Ryen & Svensson, 2015). Furthermore, it is important to note that all IV  $CIV_{QALY}$  estimates, except the log-income specification, fell within the range of  $v_Q$  estimates for Germany of €4988 to €43,115 reported by Ahlert et al. (2016), who provided the only  $v_Q$  estimates until now. A first approximation of an *opportunity cost based QALY threshold value*, or  $k_Q$ , for Germany was reported by Woods et al. (2016). Using empirical estimates of health care opportunity costs for Germany, and the relationship between GDP per capita and the value of a statistical life, they calculated a  $k_Q$  range of €19,276 to €24,374 (in 2018 euros). A recent related study by Ochalek and Lomas (2020) reported estimates of cost per DALY averted (essentially the reciprocal of a QALY gain) for Germany of €47,116 to €74,650 (in 2018 euros).

## 6.2 | Limitations and strengths of the analysis

IV-based estimates rely on restrictive assumptions relating to their unbiasedness and informational value. A valid concern is that occupational choice may be related to other unobserved confounders, such as personality traits or income preferences (Pischke & Schwandt, 2012). The use of individual fixed-effects should somewhat alleviate such concerns due to the rather stable nature of personality traits (Borghans, Duckworth, Heckman, & ter Weel, 2008), but they cannot provide complete assurance. A further assumption is that being employed in a certain industry/occupation should not have a significant, direct effect on life satisfaction, therefore violating the exclusion restriction. Supplementary Figures A3.6 and A3.7 show that, controlling for income and other confounders, this effect is not zero, but modest and mostly insignificant. One additional drawback that is rarely explicitly discussed but of great importance in the well-being valuation context, is that IV estimates only yield a local average treatment effect (Angrist et al., 1996). Using predicted labour income as an instrument, at least questions the generalisability of our IV estimates to the full, also non-working, population. Further, as we are not able to address all sources of measurement error with respect to income, the remaining upward bias in the income coefficients would imply a downward bias in the estimated  $CIV_{QALY}$  values.

In addition, income variation in industry-occupation cells predominantly consists of *positive, upward shifts in wages* (and differences therein). This is conceptually different to financial worsening events, as used by Huang et al. (2018), as these capture *income losses*.<sup>7</sup> Given income loss aversion (Boyce et al., 2013), our IV based  $CIV_{QALY}$  estimates likely represent a lower-bound.

The potential endogeneity of health (status) in life satisfaction regressions due to reverse causality (see e.g., Veenhoven, 2008 or Sabatini, 2014), which is rarely addressed in the related literature, is a further limitation. This endogeneity could be addressed by appropriate instruments or identifying health shocks which are plausibly exogenous, such as heart attacks or strokes. However, besides practical issues like data availability, it is questionable how generalisable such localized causal effects would be for the overall impact of the multi-dimensional construct of health on life satisfaction. Heterogeneity may exist both concerning the type of health shocks, but also relating to their timing within the (life-cycle) health distribution. Whether or not our estimates of the impact of health are biased upwards or downwards can therefore not be easily ascertained. In the one previous article in the related literature that addressed endogeneity directly, Brown (2015) found that the health coefficient was slightly overestimated when not instrumented. Assuming this also holds in our context, this would imply that there is an upward bias in our  $CIV_{QALY}$  values resulting from the endogeneity of health.

A more practical limitation relating to measuring health was that we had to impute SF-6D utilities for every second year to make full use of the SOEP's rich annual data. This required us to condition the sample on individuals who had at least three consecutive observations, which may have resulted in underestimating the impact of deteriorating health, since individuals are more likely to discontinue their participation in a longitudinal survey following a negative health shock.

A final limitation lies in the potential presence of double-counting as SWB enters the model twice: As an implicit consideration in the SF-6D health state valuation tasks (on which the scoring of our health measure is based on), and as a proxy for experienced utility (Equation 2). To what extent this is problematic is difficult to assess. To avoid this double counting one could use an unweighted sum score of the SF-6D levels. However, this raises the question of the appropriate anchoring. Using such a sum score, rescaled to a 0 to 1 range (expanding the number of levels of the first two SF-6D dimensions to five to not impose any weighting) lead to lower  $CIV_{QALY}$  estimates in the unimputed dataset (Supplementary Table A2.2, columns 4–5). However, when imposing the same anchor and therefore range as in the original SF-6D tariff (0.345 to 1), the OLS and IV results (€88,867 and €30,567) were much closer to the unimputed baseline estimates (€80,671 and €27,777).

It seems that not the differential weighting between the dimensions caused the larger differences, but the different anchors, that is the lowest utility. Another alternative approach entailed eliciting  $CIV$  values for different dimensions directly by regressing on all levels of the SF-6D, which did not impose any weighting. Adding up the resulting  $CIV$  values of the lowest level of all six dimensions, summed up to a cumulative value of moving from the best possible to the worst possible health state of €79,013 and €27,489, which again resembled the unimputed baseline estimate (Supplementary Table A2.2). While these sensitivity checks somewhat alleviate the concerns about double-counting, the latter revealed that 46% of the  $CIV_{QALY}$  value stemmed from the impact of mental health on life satisfaction. It is likely that the mental health dimension also plays a dominant role in our baseline calculations. Whether this in itself is problematic lies outside the scope of this paper, as it relates to a more general issue of the well-being valuation approach: is life satisfaction the best (available) proxy for experienced utility?

### 6.3 | Implications of findings

There are several practical implications of our study for future applications of the well-being valuation approach in general, and its use for estimating  $v_Q$  in particular. First, judging from the impact outliers have in the OLS specification (Table 4), subsequent applications of the approach using linear models should report on the occurrence and treatment of outliers. Secondly, given that the functional form of income had a large impact on our estimates its final specification has to be well argued and reporting results for other alternative functional forms seems warranted. The piecewise linear specification seems to be a promising alternative, given that it is more flexible and gives all income groups a proportional weight. This approach, however, comes at the price of increasing the number of variables that need to be instrumented for.

Third, the choice of utility tariffs for the health instrument matters greatly. Especially the range of the scoring algorithm has a large impact (Supplementary Table A2.2), as an imposed one unit change in health utility implies a different change in health if the range goes from 0.345 to 1 or  $-0.44$  to 1. How to overcome this issue while facilitating cross-country comparisons and how this relates to the underlying QALY concept, should further be discussed in future applications. Lacking country specific tariffs, it may be convenient to opt for a tariff whose origin



can be placed in cultural and socio-economic proximity to the country to be investigated. However, the impact of methodological peculiarities in how these tariffs were generated is relevant. It would have been interesting also to compute  $CIV_{QALY}$  estimates based on the more widely used EQ-5D health utilities and compare the implications of differences in scope and range of the health instrument used on  $CIV_{QALY}$  values. Unfortunately, EQ-5D is rarely included longitudinal surveys. Lastly, the differing values obtained when considering East and West Germany separately, or specific time periods (Table 3), also highlight the potential importance of the specific country context for  $CIV_{QALY}$  calculations.

One of the major conceptual issues discussed in our analysis, with direct relevance for the practical value of any empirically estimated  $CIV$  of health, is the health state dependence of utility. We attempted to provide indicative evidence on how health state dependence might affect estimated  $CIV_{QALY}$  values. However, it remains unclear whether empirical approaches based on self-reported (panel) data can produce reliable estimates if health state dependence is prevalent and survey participation and attrition is (partially) driven by health changes over time. We found considerable differences in the estimated  $CIV_{QALY}$  values when comparing periods of good and bad health within individuals (Table 6). As the underlying point estimates depicted substantial uncertainty, these findings should be interpreted with caution and merely as indicative evidence for the role of health state dependency in this context. The impact of this subsample of individuals on the population wide  $CIV_{QALY}$  value is likely small, as attrition is high once individuals experience bad health states, long-term or very severe health shocks. Hence, a pragmatist might argue that this issue is of theoretical interest only. We would argue, however, that this is an inherent limitation of self-reported observational data and its *ex-post* perspective in this context. Stated preference methods would allow for an explicit *ex-ante* consideration of this issue through tailored sampling strategies and survey design.

An additional conceptual concern related to health state dependence is the question of adaptation to bad health over time (Huang et al., 2018). Adaptation implies the gradual return of SWB to pre-health-shock levels despite continued (or deteriorating) bad health (Loewenstein & Ubel, 2008). This phenomenon has been documented before using the SOEP-data (Oswald & Powdthavee, 2008) and would generally decrease estimated  $CIV_{QALY}$ , as the marginal utility of health would decrease with time spent in bad health. To what extent this represents an estimation error, however, is debatable and depends on what is perceived to be the “true” impact of ill-health on well-being over time, and whether adaptation, if present, should be corrected for. The recent findings by Etilé et al. (2020), who documented a heterogeneous distribution of adaptive potential across subgroups, underline the relevance of this concern also from a normative perspective.

The previous remarks highlight avenues for future research, like investigating the causal effect of health on life satisfaction, for example using instrumental variable regressions. In addition, the approach would crucially benefit from further research into the impact of income on life satisfaction, for example using (natural) experiments. The regular inclusion of variables that represent valid instruments for income into different population panel surveys could also be beneficial for further exploring the reliability and validity of these instruments and the approach as a whole, as it would allow cross-national replications of results. Meanwhile, future applications may draw upon recent advances into the generalisability of IV-based estimates (see e.g., Mogstad et al. (2018)) to explore how these concerns can be addressed within the framework of available instruments. Further, linking survey data on individual-level SWB measures with detailed administrative records on income, health, and care consumption would also be a fruitful direction for further inquiry, resolving some of the enumerated concerns. With respect to the question of health state dependency, for example, it would be possible to determine the extent to which survey data has an inherent blind spot due to the attrition of individual following severe health shocks. In addition, such data could also be used to explore a wider range of specification choices within the general empirical strategy used, for example with respect to the choice of control variables. Here, we deliberately followed Huang et al. (2018), as the set of basic control variables they propose is available in most national panel surveys, which facilitates replications across country-contexts. However, there is ample room for extending the analysis by considering a wider set of control variables and their impact on  $CIV_{QALY}$  estimates, or even to altogether choose a different approach such as shrinkage estimators (e.g., LASSO) or matching to address endogeneity concerns around the impact of health and/or income on life satisfaction.

A final issue concerns the practical application of our  $v_Q$  estimates. If certain (health) policies/interventions in Germany were to be evaluated using a  $v_Q$  value from our study, which range from around €20,000 (IV) to €60,000 (OLS), we have to highlight the following:<sup>8</sup> Our study cannot provide a definite answer regarding which estimate is most accurate to be used in different contexts. This relates to the uncertainty surrounding these estimates and the underlying assumptions, but also to normative or distributional questions, which need to be addressed in the future (Cookson

et al., 2020). While our piecewise regression results somewhat reflect such concerns by constructing  $v_Q$  estimates using a weighted mean of the different parts of the income distribution, this is only a first, very simplistic approach. When used in a normative context, like decisions on reimbursement of technologies, explicit policy (debate and) support is required. Applied studies could use the range we provided to highlight the impact of varying  $v_Q$  estimates on their results and recommendations, keeping in mind that for specific sub-populations our  $v_Q$  estimates might not be directly applicable. In any case the selection of any specific value over another in any practical application should be transparently discussed with respect to the applied selection criteria.

## 7 | CONCLUSIONS

We demonstrated that the well-being valuation approach *can* be another useful instrument in the (health) economist's tool box for obtaining monetary equivalent valuations of health ( $v_Q$ ). Some inherent empirical and conceptual challenges of applying this approach in this context can be addressed, especially when using large-scale longitudinal data. However, other issues, like the health state dependence of the utility of consumption, will remain a threat to the validity of estimates, warranting additional research. Concurrently, alternative approaches of estimating  $v_Q$ , like stated preference studies or methods aiming at eliciting the value of a statistical life, as recently applied by Herrera-Araujo et al. (2020), provide important complementary insights, despite their conceptual differences. Also given their respective strengths and limitations, methodological diversity is desired in the ongoing endeavour of measuring the monetary equivalent value of health.

The type of  $v_Q$  estimates provided in our analysis reflect average marginal health valuations (with the caveat of being entirely based on marginal changes in health related quality of life), representative on a national level. As such, these can be applied in economic evaluations informing decision making on a societal level for publicly funded policies or interventions. Such  $v_Q$  estimates predominantly find their use by informing the cost-effectiveness threshold in the context of cost-utility analysis within health care, which aid in informing decisions on reimbursement of certain health interventions. However, estimates of the monetary value of health can also be useful in broader contexts, like cost-benefit analyses or similar approaches (Cookson et al., 2020), especially when benefits and costs of policies/interventions constitute a mix of health and non-health outcomes occurring across different sectors. Advancing methodologies aiming to estimate  $v_Q$  and providing insights into their validity can assist in informing some of the uncomfortable trade-offs that societies generally face in priority-setting both within health care but also beyond (Chilton et al., 2020).

## ACKNOWLEDGEMENTS

We would like to thank two anonymous reviewers for their insightful and constructive comments. Further we would like to thank participants and especially discussants at the Nordic Health Economists' study group meeting 2019, the EuHEA PhD conference 2019, the 2019 International Health Economics Congress and the 2019 meeting of the German Health Economics Association. We further would like to thank Dorte Gyrd-Hansen and Tom Stargardt for excellent comments on a previous version of this paper. Analogous to the functioning of a hydraulic pump, the consistent input pressure on this work, exerted by all mentioned parties, enabled us to produce continuous and sustained output (improvements). Sebastian Himmler receives funding from a Marie Skłodowska-Curie fellowship financed by the European Commission (Grant agreement No. 721402) and Jannis Stöckel receives funding from the Erasmus Initiative Smarter Choices for Better Health of the Erasmus University Rotterdam. All remaining errors are our own.

## CONFLICTS OF INTERESTS

None of the authors has any conflict of interest to declare.

## DATA AVAILABILITY STATEMENT

The SOEP scientific use file with anonymized microdata is made available at no cost to universities and research institutes for non-commercial research and teaching purposes in various data formats. To obtain data access users must sign a user contract and confidentiality agreement (<http://www.diw.de/soep-contractmanagement>). After approval of this contract the data can be downloaded from the SOEP website via a secure data transfer system. For further

instructions please see the data access information on the SOEP website ([https://www.diw.de/en/diw\\_01.c.601584.en/data\\_access.html](https://www.diw.de/en/diw_01.c.601584.en/data_access.html)).

This publication used the SOEP core v35 (1984 - 2018) data release. A replication package can be found on the Open Science Framework (<https://osf.io/b8nsz/>). This replication package includes all Stata dofiles used to obtain the presented results as well as those needed to construct the dataset used for the analyses based on the raw SOEP v35 data release.

## ORCID

Sebastian Himmler  <https://orcid.org/0000-0002-6565-5353>

Jannis Stöckel  <https://orcid.org/0000-0001-8067-7422>

Job van Exel  <https://orcid.org/0000-0002-4178-1777>

Werner B. F. Brouwer  <https://orcid.org/0000-0002-0476-8397>

## ENDNOTES

- <sup>1</sup> In the broader cost-benefit framework, this quality-adjusted life year (QALY) equivalent  $v_Q$  value can be used for transforming health gains into monetary benefits.
- <sup>2</sup> Instead, the trade-off between  $\Delta c_i$  and  $\Delta Q$  is discussed and determined in closed-door price negotiations between health authorities and the manufacturer. The methodological uncertainty around estimating  $v_Q$  has been cited as a key reason for the scepticism towards adopting more transparent threshold-based decision rules (Bundesministerium für Gesundheit, 2008).
- <sup>3</sup> The SF-12 is also used to calculate SF-6D health utilities. Component scores range from 0 (worst) to 100 (best) with a normalized mean of 50 and standard deviation of 10 (Ware et al., 1995).
- <sup>4</sup> Following Luechinger (2009) we added a constant of €1 to all incomes for the log-income specification.
- <sup>5</sup> Bayer and Juessen (2015) used only data from West Germany, possibly leading to a downward bias due to higher income levels in the West. Similarly, both Pischke (2011) and Luechinger (2009) use SOEP waves from the years before the East German SOEP sample was established in 1990 alongside waves containing samples from both former German states past 1990.
- <sup>6</sup> Huang et al. (2018) did not observe a large difference between linear and log income based estimates. However, they multiplied the ratio of income and health coefficients as in Equation (6) with the median income to obtain  $CIV_{QALY}$  (as opposed to Equation (10)).
- <sup>7</sup> Ambrosio et al. (2018) report a persistent direct effect of financial worsening (and improvement) events on life satisfaction beyond income-changes, raising concerns on the general appropriateness such events as income instruments.
- <sup>8</sup> Health care funding decisions in Germany are not based on cost utility analysis, partially because thresholds were considered to be difficult to define (Bundesministerium für Gesundheit, 2008). Finding comparable monetary estimates using a compensating income variation and stated preference studies to some extent puts this into perspective.

## REFERENCES

- Ahlert, M., Breyer, F., & Schwetmann, L. (2016). How you ask is what you get: Framing effects in willingness-to-pay for a QALY. *Social Science & Medicine*, 150, 40–48. <https://doi.org/10.1016/J.SOCSCIMED.2015.11.055>
- Ambrosio, C. D., Clark, A., & Zhu, R. (2018). *Living in the shadow of the past : Financial profiles , health and well-being living in the shadow of the past : Financial profiles, health and well-being*. Working paper. Retrieved from <http://www.iariw.org/copenhagen/dambrosio.pdf>
- Angrist, J. D., Imbens, G. W., & Rubin, D. B. (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association*, 91(434), 444–455. <https://doi.org/10.2307/2291629>
- Bayer, C., & Juessen, F. (2015). Happiness and the persistence of income shocks. *American Economic Journal: Macroeconomics*, 7(4), 160–187. <https://doi.org/10.1257/mac.20120163>
- Bobinac, A., van Exel, N. J. A., Rutten, F. F. H., & Brouwer, W. B. F. (2012). Get more, pay more? An elaborate test of construct validity of willingness to pay per qaly estimates obtained through contingent valuation. *Journal of Health Economics*, 31(1), 158–168. <https://doi.org/10.1016/j.jhealeco.2011.09.004>
- Bobinac, A., van Exel, N. J. A., Rutten, F. F. H., & Brouwer, W. B. F. (2013). Valuing QALY gains by applying a societal perspective. *Health Economics*, 22(10), 1272–1281. <https://doi.org/10.1002/hec.2879>
- Bollen, K. A., & Jackman, R. W. (1985). Regression diagnostics. *Sociological Methods & Research*, 13(4), 510–542. <https://doi.org/10.1177/0049124185013004004>
- Borghans, L., Duckworth, A. L., Heckman, J. J., & Ter Weel, B. (2008). The economics and psychology of personality traits. *Journal of Human Resources*, 43(4), 972–1059. <https://doi.org/10.3368/jhr.43.4.972>
- Boyce, C. J., Wood, A. M., Banks, J., Clark, A. E., & Brown, G. D. A. (2013). Money, well-being, and loss aversion. *Psychological Science*, 24(12), 2557–2562. Retrieved from <https://doi.org/10.1177/0956797613496436> (PMID: 24126382) <https://doi.org/10.1177/0956797613496436>
- Brazier, J. E., & Roberts, J. (2004). The estimation of a preference-based measure of health from the SF-12. *Medical Care*, 42(9), 851–859.

- Brouwer, W., van Baal, P., van Exel, J., & Versteegh, M. (2019). When is it too expensive? Cost-effectiveness thresholds and health care decision-making. *The European Journal of Health Economics*, 20(2), 175–180. <https://doi.org/10.1007/s10198-018-1000-4>
- Brown, T. T. (2015). The subjective well-being method of valuation: An application to general health status. *Health Services Research*, 50(6), 1996–2018. <https://doi.org/10.1111/1475-6773.12294>
- Bundesministerium für Gesundheit. (2008). *Stellungnahme zur Methodik der Kosten-Nutzen-Bewertung von Arzneimitteln*. Retrieved from <http://www.bmg.bund.de/>
- Cai, S., & Park, A. (2016). Permanent income and subjective well-being. *Journal of Economic Behavior & Organization*, 130, 298–319. <https://doi.org/10.1016/j.jebo.2016.07.016>
- Cameron, D., Ubels, J., & Norström, F. (2018). On what basis are medical cost-effectiveness thresholds set? Clashing opinions and an absence of data: A systematic review. *Global Health Action*, 11(1), 1447828. <https://doi.org/10.1080/16549716.2018.1447828>
- Chilton, S., Nielsen, J. S., & Wildman, J. (2020). Beyond COVID-19: How the 'dismal science' can prepare us for the future. *Health Economics*, 29, 851. <https://doi.org/10.1002/hec.4114>
- Cleemput, I., Neyt, M., Thiry, N., De Laet, C., & Leys, M. (2011). Using threshold values for cost per quality-adjusted life-year gained in healthcare decisions. *International Journal of Technology Assessment in Health Care*, 27(1), 71–76. <https://doi.org/10.1017/S0266462310001194>
- Cookson, R., Skarda, I., Cotton-Barratt, O., Adler, M., Asaria, M., & Ord, T. (2020). Quality adjusted life years based on health and consumption: A summary wellbeing measure for cross-sectoral economic evaluation. *Health Economics*, 30(1), 70–85. <https://doi.org/10.1002/hec.4177>
- Dolan, P., Kavetsos, G., Krekel, C., Mavridis, D., Metcalfe, R., Senik, C., Szymanski, S., & Ziebarth, N. R. (2019). Quantifying the intangible impact of the olympics using subjective well-being data. *Journal of Public Economics*, 177, 104043. <https://doi.org/10.1016/j.jpubeco.2019.07.002>
- Donaldson, C., & Mitton, C. (2020). Coronavirus: Where has all the health economics gone? *International Journal of Health Policy Management*, 9(11), 466–468. <https://doi.org/10.34172/ijhpm.2020.108>
- Etilé, F., Frijters, P., Johnston, D. W., & Shields, M. (2020). *Psychological resilience to major socioeconomic life events*. IZA Discussion Paper No. 13063.
- Federal Statistical Office (2020). *Consumer price index (incl. rates of change): Germany, years. Table: 61111-0001*. Retrieved from <https://www-genesis.destatis.de/genesis/online?operation=sprachwechsel&language=en>
- Ferrer-i-Carbonell, A., & van Praag, B. M. S. (2002). The subjective costs of health losses due to chronic diseases. An alternative model for monetary appraisal. *Health Economics*, 11(8), 709–722. <https://doi.org/10.1002/hec.696>
- Finkelstein, A., Luttmer, E. F. P., & Notowidigdo, M. J. (2013). What good is wealth without health? The effect of health on the marginal utility of consumption. *Journal of the European Economic Association*, 11(suppl\_1), 221–258. <https://doi.org/10.1111/j.1542-4774.2012.01101.x>
- Frey, B. S., Luechinger, S., & Stutzer, A. (2009). The life satisfaction approach to valuing public goods: The case of terrorism. *Public Choice*, 138(3–4), 317–345. <https://doi.org/10.1007/s1127-008-9361-3>
- Frijters, P., Haisken-DeNew, J. P., & Shields, M. A. (2004). Money does matter! evidence from increasing real income and life satisfaction in east Germany following reunification. *The American Economic Review*, 94(3), 730–740. <https://doi.org/10.1257/0002828041464551>
- Goebel, J., Grabka, M. M., Liebig, S., Kroh, M., Richter, D., Schröder, C., & Schupp, J. (2019). The German socio-economic panel (soep). *Jahrbucher für Nationalökonomie und Statistik*, 239(2), 345–360.
- Gyrd-Hansen, D. (2003). Willingness to pay for a qaly. *Health Economics*, 12(12), 1049–1060. <https://doi.org/10.1002/hec.799>
- Gyrd-Hansen, D., Jensen, M. L., & Kjaer, T. (2014). Framing the willingness-to-pay question: Impact on response patterns and mean willingness to pay. *Health Economics*, 23(5), 550–563. <https://doi.org/10.1002/hec.2932>
- Hagenaars, A. J., De Vos, K., Asghar Zaidi, M. (1994). Statistical Office of the European Communities. Poverty statistics in the late 1980s: Research based on micro-data. Office for Official Publications of the European Communities.
- Hariri, J. G., & Lassen, D. D. (2017). Income and outcomes. *Public Opinion Quarterly*, 81(2), 564–576. <https://doi.org/10.1093/poq/nfw044>
- Hendren, N., & Sprung-Keyser, B. (2020). A unified welfare analysis of government policies. *Quarterly Journal of Economics*, 135(3), 1209–1318.
- Herrera-Araujo, D., Hammitt, J. K., & Rheinberger, C. M. (2020). Theoretical bounds on the value of improved health. *Journal of Health Economics*, 72, 102341. <https://doi.org/10.1016/j.jhealeco.2020.102341>
- Himmler, S., van Exel, J., & Brouwer, W. (2020). Estimating the monetary value of health and capability well-being applying the well-being valuation approach. *The European Journal of Health Economics*, 21, 1235–1244. <https://doi.org/10.1007/s10198-020-01231-7>
- Howley, P. (2017). Less money or better health? Evaluating individual's willingness to make trade-offs using life satisfaction data. *Journal of Economic Behavior & Organization*, 135, 53–65. <https://doi.org/10.1016/j.jebo.2017.01.010>
- Huang, L., Frijters, P., Dalziel, K., & Clarke, P. (2018). Life satisfaction, QALYs, and the monetary value of health. *Social Science & Medicine*, 211(June), 131–136. <https://doi.org/10.1016/j.socscimed.2018.06.009>
- Jonker, M. F., Donkers, B., de Bekker-Grob, E. W., & Stolk, E. A. (2018). Advocating a paradigm shift in health-state valuations: The estimation of time-preference corrected QALY tariffs. *Value in Health*, 21(8), 993–1001. <https://doi.org/10.1016/j.jval.2018.01.016>
- Kling, C. L., Phaneuf, D. J., & Zhao, J. (2012). From Exxon to BP: Has some number become better than No number? *The Journal of Economic Perspectives*, 26(4), 3–26. <https://doi.org/10.1257/jep.26.4.3>

- Kools, L., & Knoef, M. (2019). Health and consumption preferences; estimating the health state dependence of utility using equivalence scales. *European Economic Review*, *113*, 46–62. <https://doi.org/10.1016/j.eurocorev.2018.12.007>
- Layard, R., Mayraz, G., & Nickell, S. (2008). The marginal utility of income. *Journal of Public Economics*, *92*(8–9), 1846–1857. <https://doi.org/10.1016/j.jpubeco.2008.01.007>
- Loewenstein, G., & Ubel, P. A. (2008). Hedonic adaptation and the role of decision and experience utility in public policy. *Journal of Public Economics*, *92*(8–9), 1795–1810.
- Luechinger, S. (2009). Valuing air quality using the life satisfaction approach. *The Economic Journal*, *119*, 482–515. <https://doi.org/10.1111/j.1468-0297.2008.02241.x>
- Luechinger, S., & Raschky, P. A. (2009). Valuing flood disasters using the life satisfaction approach. *Journal of Public Economics*, *93*(3–4), 620–633. <https://doi.org/10.1016/j.jpubeco.2008.10.003>
- Mason, H., Jones-Lee, M., & Donaldson, C. (2009). Modelling the monetary value of a QALY: A new approach based on UK data. *Health Economics*, *18*(8), 933–950. Retrieved from <http://doi.wiley.com/10.1002/hec.1416>
- Mcdonald, R., & Powdthavee, N. (2018). *The shadow prices of voluntary caregiving : Using panel data of well-being to estimate the cost of informal care*. IZA DISCUSSION PAPER SERIES. 11545.
- McIntosh, E. (2010). Introduction. In E. McIntosh, P. Clarke, E. Frew, & J. Louviere (Eds.), *Applied methods of cost–benefit analysis in health care* (pp. 3–18). Oxford University Press.
- McNamee, P., & Mendolia, S. (2018). Changes in health-related quality of life: A compensating income variation approach. *Applied Economics*, *51*(00), 639–650. <https://doi.org/10.1080/00036846.2018.1504160>
- Mogstad, M., Santos, A., & Torgovitsky, A. (2018). Using instrumental variables for inference about policy relevant treatment parameters. *Econometrica*, *86*(5), 1589–1619. <https://doi.org/10.3982/ECTA15463>
- Neumann, P. J., Sanders, G. D., Russell, L. B., Siegel, J. E., & Ganiats, T. G. (2016). *Cost effectiveness in health and medicine*. Oxford University Press.
- Norman, R., Cronin, P., Viney, R., King, M., Street, D., & Ratcliffe, J. (2009). International comparisons in valuing EQ-5D health states: A review and analysis. *Value in Health*, *12*(8), 1194–1200. <https://doi.org/10.1111/j.1524-4733.2009.00581.x>
- Ochalek, J., & Lomas, J. (2020). Reflecting the health opportunity costs of funding decisions within value frameworks: Initial estimates and the need for further research. *Clinical Therapeutics*, *42*(1), 44–59.e2. Retrieved from <https://doi.org/10.1016/j.clinthera.2019.12.002>
- Ólafsdóttir, T., Ásgeirsdóttir, T. L., & Norton, E. C. (2020). Valuing pain using the subjective well-being method. *Economics and Human Biology*, *37*, 100827. <https://doi.org/10.1016/j.ehb.2019.100827>
- Oswald, A. J., & Powdthavee, N. (2008). Does happiness adapt? A longitudinal study of disability with implications for economists and judges. *Journal of Public Economics*, *92*(5–6), 1061–1077. <https://doi.org/10.1016/j.jpubeco.2008.01.002>
- Phelps, C. E. (2019). A new method to determine the optimal willingness to pay in cost-effectiveness analysis. *Value in Health*, *22*(7), 785–791. Retrieved from <https://doi.org/10.1016/j.jval.2019.03.003>
- Philipson, T. J., & Jena, A. B. (2006). *Forum for health economics & policy* (Vol. 9). Who benefits from new medical technologies? Estimates of consumer and producer surpluses for HIV/AIDS drugs
- Pinto-Prades, J.-L., Sánchez-Martínez, F.-I., Corbacho, B., & Baker, R. (2014). Valuing qalys at the end of life. *Social Science & Medicine*, *113*, 5–14. <https://doi.org/10.1016/j.socscimed.2014.04.039>
- Pischke, J.-S. (2011). Money and happiness: Evidence from the industry wage structure. *NBER Working Paper 17056*.
- Pischke, J.-S., & Schwandt, H. (2012). *A cautionary note on using industry affiliation to predict income* (Tech. Rep.). National Bureau of Economic Research.
- Rousseeuw, P. J., & Leroy, A. M. (1987). *Robust regression and outlier detection* (Vol. 1). Wiley Online Library.
- Rowen, D., Azzabi Zouraq, I., Chevrou-Severac, H., & van Hout, B. (2017). International regulations and recommendations for utility data for health technology assessment. *Pharmacoeconomics*, *35*(s1), 11–19. <https://doi.org/10.1007/s40273-017-0544-y>
- Ryen, L., & Svensson, M. (2015). The willingness to pay for a quality adjusted life year: A review of the empirical literature. *Health Economics*, *24*(10), 1289–1301. <https://doi.org/10.1002/hec.3085>
- Sabatini, F. (2014). The relationship between happiness and health: Evidence from Italy. *Social Science & Medicine*, *114*, 178–187. <https://doi.org/10.1016/j.socscimed.2014.05.024>
- Socio-Economic Panel, S. (2019). *Socio-economic panel (soep), data for years 1984–2018, version 35*. SOEP.
- Soeteman, L., van Exel, J., & Bobinac, A. (2017). The impact of the design of payment scales on the willingness to pay for health gains. *The European Journal of Health Economics*, *18*(6), 743–760. <https://doi.org/10.1007/s10198-016-0825-y>
- Szymanski, P., & Olsen, J. A. (2001). Equity in health: The importance of different health streams. *Journal of Health Economics*, *20*(5), 823–834. [https://doi.org/10.1016/s0167-6296\(01\)00095-9](https://doi.org/10.1016/s0167-6296(01)00095-9)
- van den Berg, B., & Ferrer-i-Carbonell, A. (2007). Monetary valuation of informal care: The well-being valuation method. *Health Economics*, *16*(11), 1227–1244. <https://doi.org/10.1002/hec.1224>
- Vatter, J. (2020). Well-being in Germany: What explains the regional variation? No. 435. *SOEP Papers*.
- Veenhoven, R. (2008). Healthy happiness: Effects of happiness on physical health and the consequences for preventive health care. *Journal of Happiness Studies*, *9*(3), 449–469. <https://doi.org/10.1007/s10902-006-9042-1>
- Verardi, V., & Croux, C. (2009). Robust regression in Stata. *STATA Journal*, *9*(3), 439–453. <https://doi.org/10.1177/1536867X0900900306>
- Wagstaff, A. (1991). Qalys and the equity-efficiency trade-off. *Journal of Health Economics*, *10*(1), 21–41. [https://doi.org/10.1016/0167-6296\(91\)90015-F](https://doi.org/10.1016/0167-6296(91)90015-F)

- Ware, J. E., Keller, S. D., & Kosinski, M. (1995). *Sf-12: How to score the sf-12 physical and mental health summary scales*. New England Medical Center: Health Institute.
- Woods, B., Revall, P., Sculpher, M., & Claxton, K. (2016). Country-level cost-effectiveness thresholds: Initial estimates and the need for further research. *Value in Health*, 19(8), 929–935. <https://doi.org/10.1016/j.jval.2016.02.017>

## SUPPORTING INFORMATION

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

**How to cite this article:** Himmler, S., Stöckel, J., van Exel, J., & Brouwer, W. B. F. (2021). The value of health—Empirical issues when estimating the monetary value of a quality-adjusted life year based on well-being data. *Health Economics*, 30(8), 1849–1870. <https://doi.org/10.1002/hec.4279>