Comparing the Noncomparable: The Need for Equivalence Measures That Make Sense in Health-Economic Evaluations

F. Reed Johnson, PhD,[1,*] Frank I. Scott, MD, MSCE,[2] Shelby D. Reed, PhD,[1] James D. Lewis, MD, MSCE,[3] Meenakshi Bewtra, MD, MPH, PhD[3]

[1]Department of Population Health Sciences & Duke Clinical Research Institute, Duke University, Durham, NC, USA; [2]Division of Gastroenterology and Hepatology, University of Colorado Anschutz Medical Campus, Aurora, CO, USA; [3]Division of Gastroenterology, Department of Biostatistics and Epidemiology, University of Pennsylvania, PA, USA.

ABSTRACT

Background: The popularity of quality-adjusted life years (QALYs) has been resistant to concerns about validity and reliability. Utility-theoretic outcome equivalents are widely used in other areas of applied economics. Equivalence values can be derived for time, money, risk, and other metrics. These equivalence measures preserve all available information about individual preferences and are valid measures of individual welfare changes.

Objective: The objective of this study was to derive alternative generalized equivalence measures from first principles and illustrate their application in an empirical comparative-effectiveness example.

Methods: We specify a general-equilibrium model incorporating neoclassical utility functions, health production function, severity-duration preferences, and labor-market tradeoff function. The empirical implementation takes advantage of discrete-choice experiment methods that are widely accepted in other areas of applied economics and increasingly in health economics. We illustrate the practical significance of restrictive QALY assumptions using comparative-effectiveness results based on both QALYs and estimates of welfare-theoretic time-equivalent values for anti-tumor necrosis factor and prolonged corticosteroid treatments for Crohn's disease in three distinct preference classes.

Results: The QALY difference between the two treatments is 0.2 months, while time-equivalent values range between 0.5 and 1.3 months for aggregate and class-specific differences. Thus, the QALY-based analysis understates welfare-theoretic values by 60%-85%.

Conclusion: These results suggest that using disease-specific equivalence values offer a meaningful alternative to QALYs to compare global outcomes across treatments. The equivalence values approach is consistent with principles of welfare economics and offers several features not represented in QALYs, including accounting for preference nonlinearities in disease severity and duration, inclusion of preference-relevant nonclinical healthcare factors, representing preferences of clinically-relevant patient subpopulations, and including utility losses related to risk aversion.

Keywords: quality-adjusted life years, time-equivalent value, willingness to pay, stated preference, discrete-choice experiments

Introduction

Ever since Fanshel and Bush[1] and Torrance et al[2] formally proposed the idea of a quality-adjusted life year (QALY) in the early 1970s, QALYs have provided a means to combine multiple dimensions of health and duration in a single metric, thereby allowing analysts to compare the impact of medical interventions across different health conditions. With the stated goal of using cost-effectiveness analysis to allocate resources efficiently to maximize the overall health of a population, the QALY has proven to be a practical metric. Analysts and policy makers benefit from its intuitive appeal and flexibility, employing QALYs to combine segments of time spent in various health states by weighting the value of health-related quality of life experienced during each segment on a scale from 0 to 1, where 0 represents dead and 1 represents full health, and summing across time.

For those wishing to make a claim to theoretical validity for the QALY, von Neumann-Morgenstern's (vNM) cardinal-utility theory provided a normative behavior model[3] when utility values for health states are derived from standard-gamble (SG) methods and

combined with probabilities that corresponding health states are realized. Nevertheless, use of vNM cardinal utility to construct a single index for complex relationships among dynamic health states requires 3 assumptions[4,5]:

- utility independence between life years and health status
- constant proportional tradeoff
- risk neutrality on life years

The assumption of utility independence between time and health status requires that utility weights for a given health state remain the same regardless of the amount of time spent in a health state and vice versa. The assumption of constant proportional tradeoff means that preference values for health states are unrelated to one's remaining life expectancy. Risk neutrality on life years means that an uncertain lifetime of given expected duration is equally preferred to a certain lifetime of the same duration. We refer to these assumptions as QALY assumptions throughout the remainder of the paper.

Empirical support for these QALY assumptions is weak. Bleichrodt and Johannesson[6] found that constant proportional tradeoff held approximately in their student sample, but the evidence was much weaker for utility independence. Only about 10% of participants' responses satisfied both requirements. Bala et al[7] and Franic et al[8] rejected utility independence in more representative samples.

The assumption of risk neutrality in the construction of QALYs allows analysts to convert ex post, or realized utility, into ex ante, or expected utility, by simply multiplying outcome utilities by their probability of occurrence. In contrast, neoclassical utility theory generally assumes that people are risk averse. Risk aversion means that utility functions are not linear, but instead, are subject to diminishing marginal utility. As a result, when faced with a choice between 2 uncertain outcomes with the same expected value, people would prefer the more certain alternative. Lakdawalla et al[9] found that patients with chronic diseases are risk averse. In contrast, Shafrin et al[10] and Lakdawalla et al[11] suggest terminal patients express a "value of hope" that appears consistent with risk-seeking preferences, hence linearity is a very strict requirement that generally is inconsistent with economic theory and observed behavior.[12,13]

In addition to empirical deviations from vNM principles, numerous commentators have identified other limitations with QALYs. One pertains to potential bias against older individuals and individuals with disabilities since the former have limited remaining life expectancy and the latter owing to concerns that their survival time may count less than a person without a disability even when other health indicators are equivalent. The German Institute for Quality and Efficiency in Health Care explicitly rejects QALYs and any form of comparative assessment across disease areas. The argument, which is also embedded in the dignity clause of article 1 of the German constitution, is that in the equity principle embedded in QALYs, equal treatment of equals (horizontal equity) is less important than positive discrimination across individuals and/or disease areas.[14] The International Society for Pharmacoeconomics and Outcomes Research Special Task Force on Value Frameworks also pointed out that QALY-based cost-effectiveness analyses fail to account for concepts such as value of hope, value of knowing, and differences in treatment processes. Bleichrodt et al[15] found that health-state valuations also depend on individual health status and expectations.

In addition to various conceptual and empirical problems with QALY measures, measuring health utility for acute conditions has proven difficult with either SG or time-tradeoff (TTO) methods. Respondents may find it difficult to compare risks of death or

reductions in longevity with self-limiting acute conditions as headache, nausea, and drowsiness that are common side effects of drug therapy. Researchers have employed various ad hoc procedures such as *chaining* to derive health-state utility values for acute conditions.[16–18]

These concerns motivate developing alternative metrics that broaden the scope of health valuation for use in comparative-effectiveness and cost-effectiveness research. Several authors have explored the consequences of including additional arguments in utility functions. Pliskin et al[4] derived risk-adjusted QALYs that accommodate risk-averse preferences. Risk-adjusted QALYs have rarely been used in empirical work, in part because it is not clear how to implement the model if quality varies over time.[19] Hammitt,[20] Dolan and Edlin,[21] Bleichrodt and Quiggin,[22] and Klose[23] investigated the consequences of incorporating wealth or non-health consumption in utility functions. Klose identified the following 3 additional requirements for such augmented utility functions to be consistent with unbiased QALY measures:

- Health is constant over time.
- Health has no influence on the marginal utility of wealth.
- Health and wealth are mutually utility independent.

These additional assumptions also are very strong and not likely to hold empirically. Moreover, they are inconsistent with Grossman's[24] influential model of the demand for health. Grossman's model specifies utility as a function of market goods and healthy days. Healthy days depend on the individual's stock of health, which can fall below the desired level because of disease. The stock of health can be replenished through a household-production process that requires health-service inputs, time, and other inputs. The demand for such inputs is derived from the demand for health.

Mehrez and Gafni[25–27] and Gafni and Birch[28] offered healthy-year equivalents (HYEs) as an early alternative to QALYs that requires fewer restrictive assumptions. The authors defined HYEs as the number of healthy years that are equivalent to a lifetime health profile. Thus, there is a unique HYE measure for each possible pattern of the timing and duration of possible periods of ill health. A lively debate ensued on the merits of the HYE approach in the mid-1990s.[27,29–32] Bleichrodt[33] subsequently concluded that "the HYE approach is theoretically sounder than the QALY approach."[33 (p30)] Nevertheless, the increase in theoretical soundness comes at the price of a significant decrease in practical feasibility because of difficulties in generalizing or transferring results of an HYE study to outcomes that differ in timing, duration, or sequence. Rittenhouse[34] also found that HYEs were a noisy alternative to conventional TTO values.

In this article, we derive a conceptual framework that allows scaling changes in well-being for equivalent changes in any continuous variable in a neoclassical (NC) utility function. Time equivalents are one case of such generalized equivalence measures. We offer an alternative approach to account for health outcomes that align with the theoretical tradition of Grossman[24] and neoclassical microeconomics in which equivalence measures correspond to individual welfare changes. We describe a welfare-theoretic conceptual framework for defining health utility and offer a practical, off-the-shelf method for estimating equivalents, using time, money, or safety as metrics for gains in health. The equivalence metrics are empirically quantified using discrete-choice experiment (DCE) methods. We illustrate the practical application of this approach and compare gains in health as measured by time-equivalents to QALYs.

In the following section, we derive health-equivalence measures. These measures are strongly linked to QALYs and HYEs —or more generally healthy-time equivalents—derived using TTO methods because we can scale welfare changes in time equivalents. Nevertheless, we do not require participants to trade off health improvements for mortality risks or reduced longevity as SG and TTO preference-elicitation tasks require. Thus, the welfare-theoretic measures are equally applicable to both acute and chronic conditions.

Utility-Maximizing Equilibrium

Welfare comparisons in conventional applied economics rely on a single reference point, usually status-quo utility, together with information on rankings of changes relative to the reference point. For example, we may want to know whether type 2 diabetes patients' perceived benefits of interventions to improve glucose control are greater than their associated cost and inconvenience. An ordinal utility scale has no meaning in absolute terms; it merely provides a way to compare relative changes in utility. Thus, an improvement in current HbA1c glucose level from 7.5 to 7 could have a utility difference of 4 units, whereas an improvement from 7.5 to 6.5 could have a utility difference of 6 units. We can only say that a difference of 6 utility units is 50% greater than a difference of 4 utility units.

For welfare measurement in most areas of applied economics, the conventional numeraire is disposable income available for consumption of all other goods and services, denominated in dollars. Suppose we knew that the mean marginal utility of a dollar of income (or reduction in cost) was 0.025 utility units. The marginal utility of income is likely greater for poorer people than for richer people, so we use the mean value for the beneficiary population. We then can rescale nominal utility differences using the "exchange rate" between dollars and utility. The money-equivalent value of a 4-unity change in utility is 4/0.025 = $160, and the value of a 6-unit change is 6/0.025 = $240. The second value is still 50% larger than the first, but rescaling in monetary units gives intuitive meaning to the utility difference as well as facilitating comparisons with dollar-denominated costs. Note that scaling utility differences in money equivalent does not make the utility scale cardinal. Although $50 is twice as much money as $25, $50 does not yield twice as much utility. A classic reference on the economic theory of cost-benefit analysis is Mishan.[35]

The dependence of money-equivalent values on the distribution of ability to pay is unacceptable to many people. We demonstrate that the same conceptual framework can be used to derive a corresponding time-equivalence preference measure. We specify a general, 1-period NC utility function as follows:

$$U = U\big[C, \ H(T_s), T_w; \overline{T}, Z_h, Z_i\big] \qquad \text{Equation 1}$$

where U is an arbitrarily scaled ordinal-utility index that is invariant with respect to any monotonic transformation. Prieto and Sacristan[36] showed that QALYs are not invariant with respect to linear transforms of utility index and discussed various consequences of this limitation. C is consumption in some clinically relevant treatment period $\overline{T}$.

For our purposes, we define the length of period $\overline{T}$ as the maximum possible duration of a spell of ill health given the individual's insured health costs. $\overline{T}$ can range from as little as an hour or less to remaining expected lifetime depending on the problem at hand. $H(T_s)$ represents the health-related utility index of a spell of illness given severity s and duration T_s. T_w is

work time. $\overline{T} - T_s$ is the amount of time spent in normal health during the period $\overline{T}$ and normal health time minus T_w is leisure time (T_l).

Z_h is a vector of healthcare process or context factors. Such factors often are important determinants of patient satisfaction related to alternative healthcare interventions. Some examples in various healthcare settings include privacy of HIV test results, dosage frequency for attention-deficit medications, diagnostic information such as genetic testing that may not affect therapeutic decisions, and differential convenience of devices such as insulin syringes and pens. Individuals also may have preferences over the features of regulatory interventions such as technology restrictions, standard-setting, or financial incentives for achieving public-health benefits. Z_i includes individual-specific demographic, health-history, attitude, and other personal-characteristic variables.

We assume individuals maximize utility over the planning period $\overline{T}$ subject to time and income constraints.

$$\overline{T} = T_l + T_w + T_s \qquad \text{Equation 2}$$

$$p_w \cdot T_w \geq p_c \cdot C + C(H) \qquad \text{Equation 3}$$

The time constraint, equation 2, assumes that the individual can work or engage in leisure activities only when healthy. Equation 3 is the income constraint. C(H) is the health-production transformation function between consumption and health indicating how much consumption must be foregone for a given level of health. Total expenditures on consumption and health must be less than total income. C^o is equal to the sum of earned income in dollars ($p_w \cdot T_w$), where p_w is the wage rate and T_w is work time, and wealth plus insured income. We normalize average consumption price p_c to equal 1.

We assume the health index exhibits diminishing marginal effects of duration of ill health and that the transformation function between C and H exhibits diminishing marginal productivity of expenditures on healthcare. Finally, we assume that utility increases at a decreasing rate with C and H. Equations 4 to 6 are the mathematical representations of these assumptions.

$$\frac{\partial^2 H}{\partial T^2} < 0 \qquad \text{Equation 4}$$

$$\frac{\partial^2 H}{\partial C^2} < 0 \qquad \text{Equation 5}$$

$$\frac{\partial U}{\partial C} < 0, \quad \frac{\partial^2 U}{\partial C} < 0, \quad \frac{\partial U}{\partial H} < 0, \quad \frac{\partial^2 U}{\partial H} < 0 \qquad \text{Equation 6}$$

Figure 1A shows tradeoffs in 4 quadrants. The northeast quadrant indicates the health-production transformation function between consumption and health, where the amount of free or insured health care is indicated by the vertical segment. Under our assumptions, after the insured amount of H, individuals must give up consumption spending to obtain healthcare. The function is nonlinear because the marginal effect of increased spending decreases as spending increases.

The northwest quadrant indicates the effect of illness duration on H. The nonlinearity of $H(T_s)$ reflects the assumption of diminishing marginal disutility of duration. As a result of adaptation, additional amounts of time with an illness of constant severity decreases health at a decreasing rate. Note that QALY assumptions require that this relationship be linear, with slope equal to the

Figure 1. (A) Consumption, health, and time tradeoffs. (B) Consumption, health, and time equilibrium. (C) Time and money equivalent values.
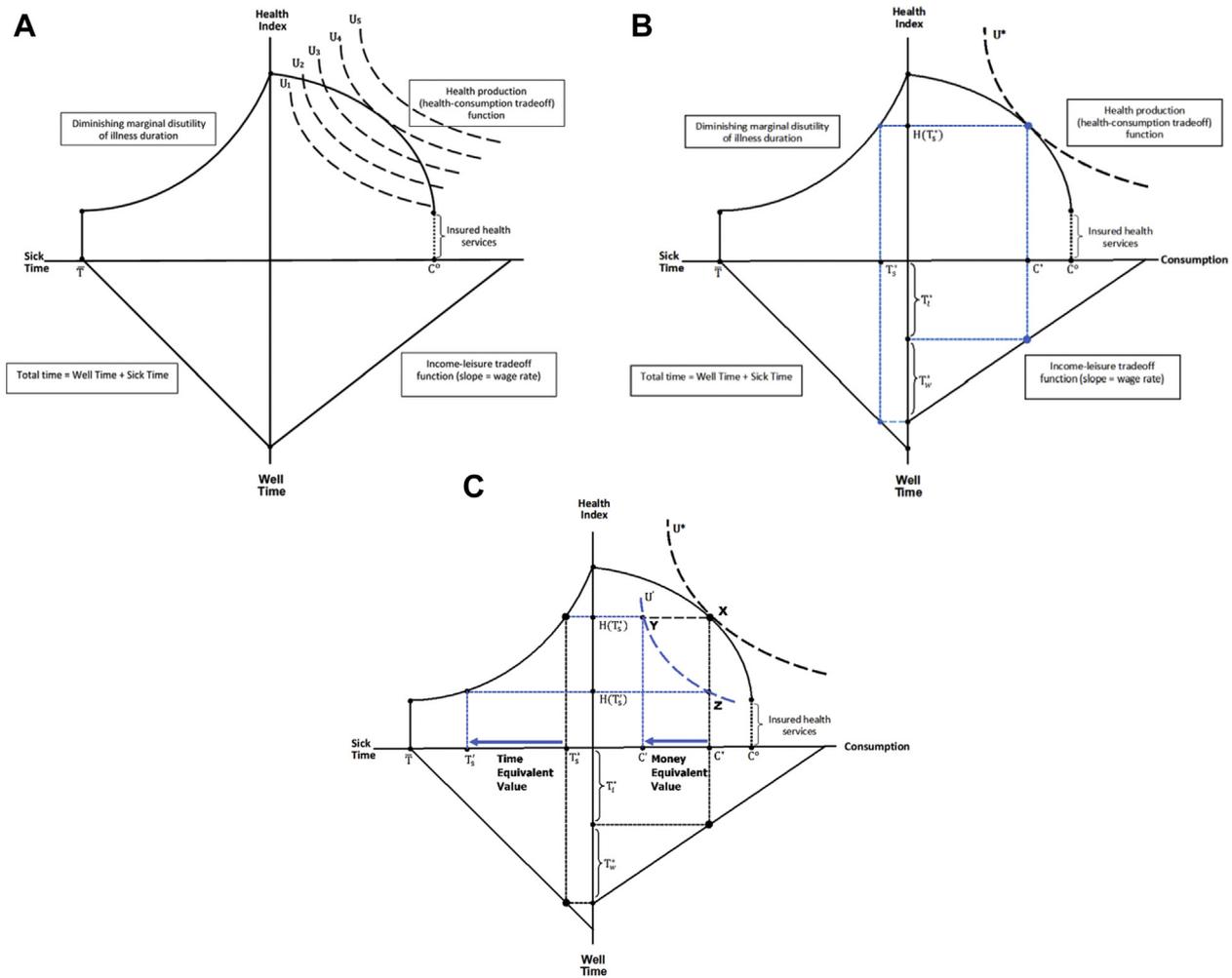
constant utility weight for condition s minus 1. The southwest quadrant enforces the adding-up constraint. One more time unit of sick time is 1 less time unit of well time available for work and leisure.

Finally, the southeast quadrant represents the income-leisure tradeoff, where the slope of the constraint is the wage rate p_w. In this quadrant, we are able to see how well time is split between work and leisure. For simplicity, we assume consumption depends only on earned income. Note that consumption will be less than potential income (indicated by the intersection of the income-leisure line with the horizontal axis) if individuals decide to allocate some of their well time to leisure instead of work. We also assume that illness duration does not affect the productivity of work time, so illnesses shorten work time, but do not change the output per hours of work time. This generalization and others could be accommodated in the model but could not be represented in 4-quadrant space. We retain the graphical representation to illustrate general equilibrium under these simplifying assumptions.

Figure 1A also shows a family of constant-utility curves U_i. Each curve indicates combinations of consumption and health that have the same utility level, with utility increasing if either C

increases, $H(T_s)$ increases, or both. Under the assumptions in equation 6, the slope of each curve decreases as consumption increases and health decreases, where the slope is

$$\frac{dH(T_s)}{dC} = -\frac{\partial U}{\partial H(T_s)} \bigg/ \frac{\partial U}{\partial C} \qquad \text{Equation 7}$$

In Figure 1B, U* is the highest utility level obtainable, given the technically feasible combinations of C and $H(T_s)$. General equilibrium among time, health, and consumption allocations is obtained at $U^* = U[C^*, H(T_s^*), T_w^*; \overline{T}, Z_h, Z_i]$. Choosing the combination C* and $H(T_s^*)$ requires simultaneous optimization in all 4 quadrants. $H(T_s^*)$ corresponds to illness duration T_s^* in the northwest quadrant. Likewise, C* requires working time equal to T_w^* in the southeast quadrant leaving T_l^* available for leisure after deducting sick time T_s^* in the southwest quadrant. Note that the positions of the curves in the northwest and northeast quadrants depend on the scale of the health index and on the type of ill health represented in the northwest quadrant.

We now are equipped to define the money-equivalent value (MEV) and time-equivalent value (TEV) of an illness that decreases utility from U* to U' in Figure 1C. The MEV is the reduced

consumption, holding health constant, that would move the individual to U'. The TEV is the reduced time in normal health, holding consumption constant, that would move the individual to U*. These values are shown in equations 8 and 9:

$$U\left[C^{*}-MEV,\ H(T_{s}^{*}),\ T_{w}^{*};\overline{T},\ Z_{h},\ Z_{i}\right] = U^{'} = U\left[C^{'},\ H\left(T_{s}^{'}\right),\ T_{w}^{'};\overline{T},\ Z_{h},\ Z_{i}\right]$$

<div align="right">Equation 8</div>

$$U\left[C^{*},\ H(T_{s}^{*}-TEV),\ T_{w}^{*};\overline{T},\ Z_{h},\ Z_{i}\right] = U^{'} = U\left[C^{'},\ H\left(T_{s}^{'}\right),\ T_{w}^{'};\overline{T},\ Z_{h},\ Z_{i}\right]$$

<div align="right">Equation 9</div>

In Figure 1C, hold illness duration constant at $H(T_{s}^{*})$, point X, and decrease consumption to decrease utility from U* to U' at point Y. The decrease C'-C* is the MEV of the utility difference between U* and U'. Alternatively, hold consumption constant at C*, point X, and reduce sick time to decrease utility from U* to U' at Z. The reduction in health from $H(T_{s}^{*})$ to $H_{s}(T_{s}^{'})$ increases illness duration from T_{s}^{*} to $T_{s}^{'}$, giving the healthy TEV of the utility difference between U* and U' as $T_{s}^{'}-T_{s}^{*}$.

Both MEV and TEV measures preserve all information about preferences contained in the underlying utility function, including income, process or context arguments, and individual characteristics. Scaling benefits in money equivalents facilitates direct comparisons to costs, whereas scaling benefits in time equivalents avoids objections to monetizing health outcomes. In principle, we could derive analogous expressions to equations 8 and 9 for any argument in the utility function. For example, we could derive risk-equivalent values, or maximum-acceptable risk, using the probability of a treatment-related adverse event as the numeraire.[37]

The MEV is often referred to as *willingness to pay*, an unfortunate term that has led to much misunderstanding. In addition, monetary values for health often have been estimated using the willingness to pay method or, more correctly, the contingent-valuation method. The contingent-valuation method was developed in the 1970s and 1980s for obtaining nonmarket values for environmental services using a modified bidding-game approach. As a result of well-documented problems with this approach, it has largely been replaced by DCE methods in environmental economics and other applied-economics research.[38] For a discussion of external validation of DCE methods, see Lancsar and Swait.[39]

Productivity costs sometimes are included in comparative-effectiveness studies by multiplying lost work time by the average wage. Nevertheless, multiplying T_{s}^{*} by the slope of the income-leisure function is not a conceptually correct welfare measure for the difference between U° and U*, even if the average wage were an accurate measure of the value of the marginal product of labor. (See Appendix B in Supplemental Materials found at https://doi.org/10.1016/j.jval.2019.03.011).

Relation of TEV to Conventional QALY Measures

Despite various assumptions of convenience that facilitate drawing the figure in 2-dimensional space, this model has several useful features. First, we impose no special requirements on utility functions. Also, unlike conventional QALY methods, we do not require utility independence between health status and time, constant and proportional tradeoffs, or risk neutrality. We also do not have to assume that the marginal utility of consumption is independent of health (or vice versa). We allow health to influence welfare both directly and indirectly via realistic time constraints. Furthermore, time durations in TEV estimates need not be quantified in years. Neither do we need to anchor utility on perfect health and death (or best-imaginable and worst-imaginable conditions). Rather, we identify equivalent values for clinically relevant health changes and illness durations.

In principle, nonmarket health values could be derived indirectly from observed behavior in this framework. For example, in environmental economics, the MEV of reducing health risks from pollution has been measured using compensating labor-market wage differentials in risky occupations, after controlling for other job characteristics.[40] Nevertheless, we generally are unable to observe the relevant money or time tradeoffs in health because of institutional constraints. For example, in highly insured health systems, the insured segment of the consumption-health tradeoff may be long enough that the duration of ill health may not reduce consumption. Also, we may be interested in evaluating new interventions for which behavioral data do not exist. As a practical matter, using this conceptual framework, we must rely on stated-preference methods to simulate tradeoff decisions to derive the relevant equivalent values. Johnson et al[41] reported TEVs for reduction in vasomotor symptoms and Hauber et al[42] reported TEVs for rotavirus vaccinations.

In the next section, we compare QALY values to TEVs obtained using this conceptual framework from a recent comparative-effectiveness study of Crohn's disease. Details on the DCE instrument and event-simulation model can be found in Lewis et al[43] and Scott et al[44] (also see Appendix A in Supplemental Materials found at https://doi.org/10.1016/j.jval.2019.03.011.)

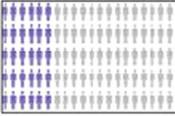## The Empirical Utility Function

Respondents with Crohn's disease evaluated pairs of treatment profiles indicating time in remission and time with symptoms with given severity over a 12-month treatment period. Additional treatment features included required time on corticosteroids, risk of serious infection, risk of lymphoma, and risk of bowel-resection surgery. Figure 2 is an example DCE choice question. Efficacy was treated as deterministic. We limited the number of probabilistic attributes to limit the cognitive burden of the preference-elicitation questions. The DCE preference-elicitation question involves tradeoffs involving severity durations. In contrast to conventional TTO studies that trade off a fixed duration with a given symptom severity and reductions in longevity, the DCE elicits tradeoffs of severity and remission durations over a clinically relevant treatment period (12 months) and clinically relevant treatment risks.

The form of the stated-preference NC utility function for specified hypothetical combinations of symptom duration, steroid duration, serious adverse-event risks, and surgery risk is:

$$U_{ij} = \sum_{t_{sev}}\sum_{sev}I_{j1}\beta_{t_{sev},sev} + \sum_{t_{ster}}I_{j2}\beta_{t_{ster}} + \sum_{p_{infect}}I_{j3}\beta_{p_{infect}} + \sum_{p_{lymph}}I_{j4}\beta_{p_{canc}} + \sum_{p_{surg}}I_{j5}\beta_{p_{surg}}$$

<div align="right">Equation 10</div>

Figure 2. Example DCE choice question, Crohn's disease treatment preferences.

where U_{ij} is the utility for individual i, treatment alternative j. I_{jk} is a 0/1 variable indicating which level appeared for that alternative. Symptom severity levels were mild, moderate, and severe; severity durations ranged from 0 to 12 months; serious infection ranged from 0% to 30%; and cancer and surgery risks ranged from 0% to 8%. Further details on the study are provided in Lewis et al[43] (also see Appendix A in the Supplemental Materials found at https://doi.org/10.1016/j.jval.2019.03.011).

Selection of the feasible subset of a potentially large list of potential attributes in a DCE study design requires careful qualitative research during survey development.[45] It is possible that it would not be possible to include all attributes that might affect decision making. Nevertheless, our studies show that careful selection of 6 to 8 attributes often captures the most salient considerations and that even a list this short typically contains attributes that do not significantly affect choices for all preference classes.

Utility-Parameter Estimates and TEV Estimates

Aggregate-sample preference-parameter estimates are shown in Figure 3. The NC utility scale on the vertical axis is arbitrary. Only relative changes are relevant. A decrease in severe symptoms from 8 months to 4 months results in a utility increase from -2.97 to -1.42 utils, an improvement of 1.55. A reduction in moderate symptoms from 12 months to 0 increases utility by 1.61. The rate of substitution or exchange rate over this interval thus is $\frac{1.55}{4} \div \frac{1.61}{12} = 2.9$ moderate months per severe month. Thus, the moderate-symptom TEV for 4 months of severe symptoms is $4 \times 2.9 = 11.6$ months. Similarly, a 30% increase in serious-infection risk reduces utility by 1.52, about the same as an increase in serious symptoms from 4 months to 8 months. Scaling this loss in equivalent-risk units yields a maximum acceptable risk of 30%. Any treatment-related risk less than 30% for the indicated

improvement in serious-symptom duration would yield a positive net utility benefit of treatment.

A more direct analog to QALYs would be normal-health or remission TEV. Because of the 12-month adding-up constraint, decreases in symptom time are perfectly inversely correlated with increases in remission time. Thus, remission utility simply is the utility gain for reducing symptom duration. Nevertheless, that value depends on the specific symptom-severity time being decreased. We thus normalize the utility gain across all symptom-severity durations by using the mean utility gain per month reduction across all changes that were evaluated in the DCE. The mean marginal utility of a 1-month reduction in symptom time was 0.17. Scaling the utility gain from reducing serious symptoms from 8 months to 4 months accordingly yields a remission TEV benefit of 1.55/0.17 = 9.1 months. Note that this is not perfect health, but rather the clinically relevant health Crohn's patients experience when their symptoms are in remission.

The NC utility is measured ex ante, that is, before health outcomes are realized. Hence, even though only small fractions of patients experience adverse events, all patients who accept a treatment experience a utility loss from bearing the risk of an adverse event. This loss generally is greater than the probability of the event multiplied by the realized utility loss, which is the risk-neutrality assumption required by conventional QALYs. We obtained TEV estimates for such risk exposures. We also estimated a latent-class model that revealed 3 classes with similar preference patterns: respondents who were relatively more concerned about treatment efficacy, respondents who were relatively more concerned about avoiding treatment risks, and respondents who were concerned about avoiding steroids.

Unlike most risk-attitude studies, we quantify risk preferences in the context of other treatment attributes. We find that many respondents will tolerate risk if offered significant improvements in health outcomes. On the other hand, some respondents were

Figure 3. NC utility estimates, Crohn's disease treatments (95% confidence intervals).
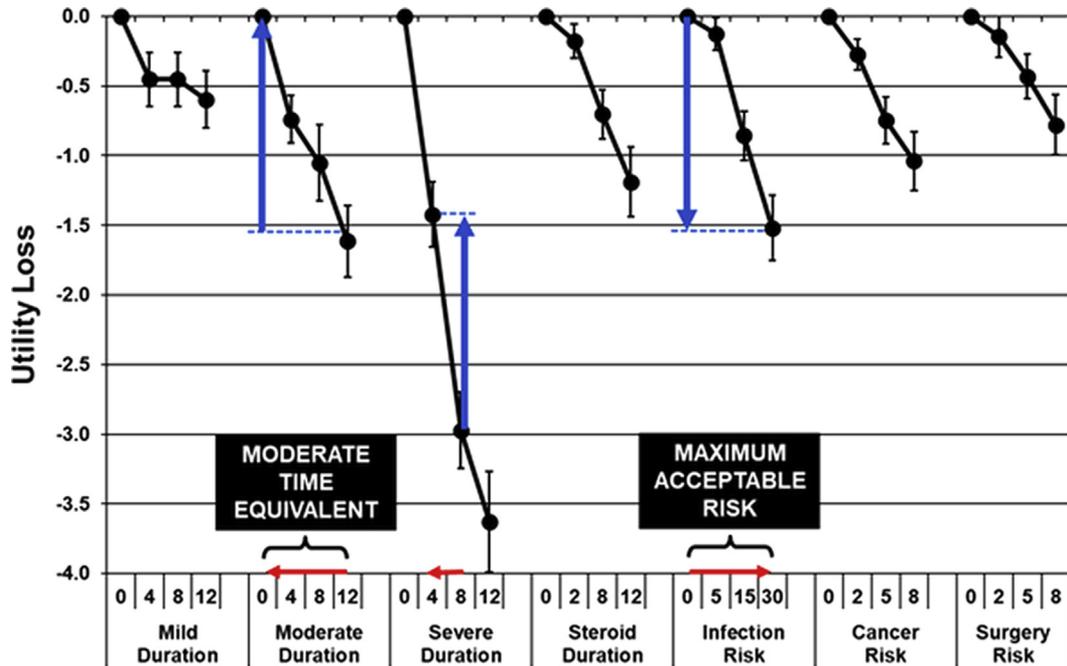

unwilling to accept any significant risk, no matter how much benefit was offered. One advantage of DCE methods is that we can observe heterogeneity in risk preferences within a sample. Figure 4 summarizes results for treatments with a 2% lymphoma risk for the aggregate sample and each preference class. Exposure to 2% lymphoma risk is equivalent to losing 7.2 months of

remission time. That loss is about 5× that of the overall sample TEV or the steroid-averse latent class.

QALY and TEV Measures of Comparative Effectiveness

Although it would be possible to calculate incremental cost-effectiveness ratios for QALY and TEV measures, we report only

Figure 4. Remission time-equivalent values, 2% treatment risk of lymphoma.
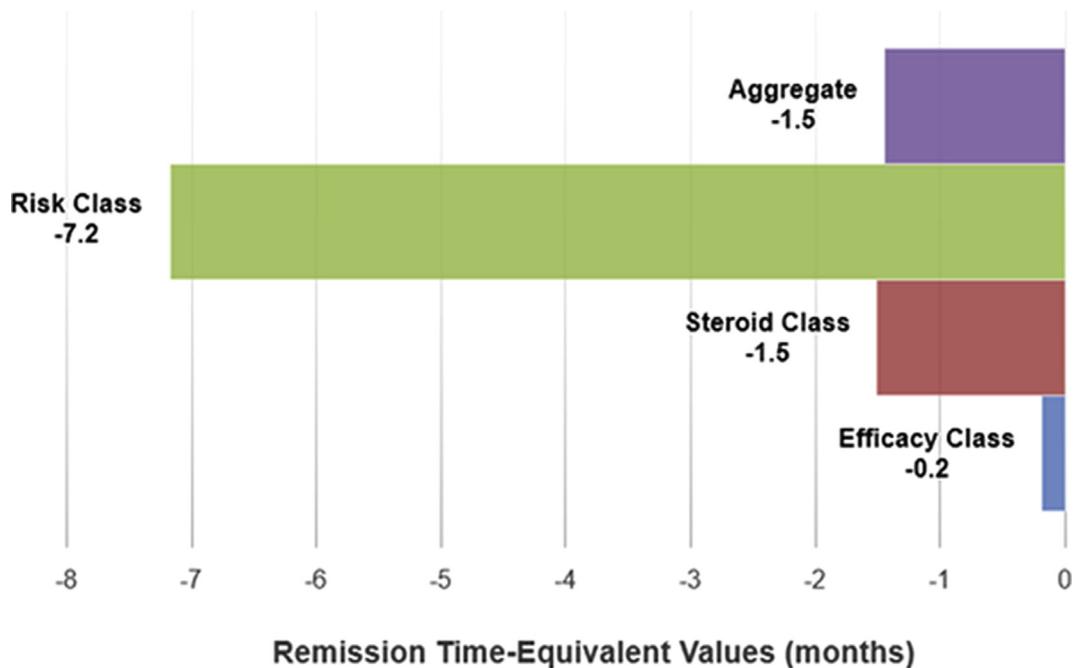
comparative effectiveness using the 2 approaches. We constructed an event-simulation model to compare corticosteroid treatment versus anti-tumor necrosis factor treatment.[46] The model was populated with QALYs based on published Crohn's health-state utilities and with severity-duration remission TEVs using both aggregate and class-specific estimates. Table 1 summarizes the results of this analysis. The QALY difference is 0.2 months, whereas TEVs range between 0.5 and 1.3 months for aggregate and class-specific differences. Thus, the QALY-based analysis understates welfare-theoretic values by 60% to 85%.

Discussion

Although we have formalized the conceptual framework for readers who are not familiar with it, we have adhered to established conventions for implementing the NC with DCE methods based on the Nobel prize-winning work of McFadden.[47] These conventions have been extensively applied and tested in every area of applied economics except health.[38] Nevertheless, there now is guidance from the US Food and Drug Administration on using patient-preference data to weight the relative importance of benefits and risks for regulatory reviews of medical devices[48] and rapidly expanding literature applying these methods in a wide range of therapeutic areas.[49] In addition, DCE methods are recommended for estimating utility weights for multi-attribute health status instruments, such as the EuroQol 5-dimensional instrument, for ultimate use in cost-effectiveness analysis.[50]

The strong assumptions underlying QALYs may give decision makers the impression that health improvements can be measured with a ruler-like metric. Nevertheless, utility functions that describe actual preferences are much more complicated. They are nonlinear and include multiple factors that describe individual preferences over alternative states of the world. Nevertheless, conventional QALY assumptions are attractive because they make empirical estimation tractable. We have shown that well-established empirical methods are available that do not require such restrictions and facilitate empirical tests of potential biases introduced by such simplifying assumptions. Both the theoretical and empirical examples indicate that distortions from QALY assumptions can be significant.

In every area of applied economics other than health economics, NC utility measures are widely accepted as a conceptually valid and powerful tool for identifying interventions that improve social welfare. In contrast, conventional QALY assumptions effectively break the link between NC utility and welfare-theoretic preferences. Value equivalents derived from NC-utility concepts preserve the link between changes in health-related utility and

changes in individual welfare in 2 ways. First, value equivalents such as MEVs and TEVs allow nonlinearity in time and heterogeneity across patients. Second, equivalent values can incorporate various health-related factors that affect individual welfare. Ignoring these additional factors undermines the relevance of QALYs for allocating scarce healthcare resources among different groups of patients and among different kinds of interventions. Our approach derives equivalent values from a common set of NC utility concepts. Using DCE utility estimates, we have calculated TEVs and maximum acceptable risks from an empirical NC utility function that includes severity duration as an argument. These measures merely are alternative ways of scaling NC utility differences for time and risk.

These results have important implications for using the QALY metric as an outcome measure in cost-utility analyses. Although it is not possible to predict the direction of bias relative to welfare-theoretic TEVs in any particular case, nonlinearities in underlying preferences could lead decision makers to misallocate healthcare resources away from short-term gains where correctly measured marginal benefits are high toward long-term gains where marginal benefits are lower. Nonlinearities similarly could bias incremental cost-effectiveness ratios in favor of more severe outcomes with low marginal benefits relative to less severe outcomes with higher marginal benefits. Including nonclinical but utility-relevant factors could alter the relative rankings of interventions if such factors vary strongly among outcomes or over time. Our results indicate that a single summary measure, whether QALY or TEV, could misrepresent the preferences of clinically relevant patient groups. Finally, treatment decisions and regulatory evaluations occur ex ante before patient outcomes are realized. Neglect of utility losses related to risk aversion thus understates the value of safety and could lead to inaccurate assessments of benefit-risk tradeoffs.

We have obtained estimates based on patient preferences, consistent with the standard welfare-theoretic framework. Nevertheless, health-state utilities elicited from the general public are recommended for use in cost-effectiveness analyses by many reimbursement agencies around the world on the basis that most healthcare expenditures are financed by taxpayers. In addition, health-state utility researchers increasingly are applying DCE methods to obtain preference estimates. Although DCE methods allow interactions between health states and durations, other restrictive assumptions are maintained. Moreover, practical applications rely on the availability of published values. In the case of the inflammatory bowel disease utilities we used in our comparative-effectiveness analysis, the only available estimates were published in 1997.

Our approach is formulated for condition-specific resource-allocation treatment decisions or for use in developing practice guidelines. We view this feature as a strength, particularly for decision making in the United States, where most within-patient and between-patient decisions are made within a therapeutic area rather than across therapeutic areas. Our time-equivalents approach allows one to incorporate condition-specific issues, such as duration of corticosteroids and risk of treatment-associated cancer, in our application to Crohn's disease. We believe that a condition-specific approach may help to address some of the resistance to using QALYs among patient advocates and other stakeholders in the United States by offering a patient-centric alternative. We understand our condition-specific approach may not be helpful to health systems seeking to allocate resources across all conditions. Those systems will continue to use the QALY. Nevertheless, even for those systems, our approach provides complementary information that can be

Table 1. Comparative effectiveness analysis, anti-TNF vs corticosteroid treatments (months)

Analysis	Anti-TNF	Prolonged corticosteroids	Incremental difference (95% CI)
QALYs	9.7	9.6	0.2
Remission TEVs			
Overall	5.3	4.5	0.8 (0.5-1.1)
Efficacy class	1.3	0.1	1.3 (0.8-1.7)
Risk class	6.9	6.4	0.6 (0.4-0.8)
Steroid class	7.8	7.3	0.5 (0.3-0.7)

CI indicates confidence interval; QALY, quality-adjusted life year; TNF, tumor necrosis factor; TEV, time-equivalent value.

compared with a QALY-based approach. It can be used to identify when treatment comparisons have concordant or conflicting findings between QALY-based and time-equivalents frameworks and may provide an indicator of when the QALY may be failing to incorporate outcomes that are important to patients.

The plausibility of any theoretical or empirical model turns on the degree to which simplifying assumptions neglect essential aspects of reality. We argue that NC utility-based methods have a better claim on validity than health-state utility-based methods. Nevertheless, we do not claim that our empirical example is a final solution to the problem of comparing noncomparables. This approach offers promise as an improvement over existing methods, but there is much more research to be done in refining both the conceptual framework and corresponding empirical applications.

Acknowledgments

Supplemental Material

Supplementary data associated with this article can be found in the online version at https://doi.org/10.1016/j.jval.2019.03.011.

REFERENCES

1. Fanshel S, Bush JW. A health-status index and its application to health-services outcomes. *Operations Res.* 1970;18(6):1021–1066.
2. Torrance GW, Thomas WH, Sackett DL. A utility maximization model for evaluation of health care programs. *Health Services Res.* 1972;7(2):118–133.
3. Von Neumann J, Morgenstern O. *Theory of Games and Economic Behavior.* Princeton, NJ: Princeton University Press; 1944.
4. Pliskin J, Shepard DS, Weinstein MC. Utility functions for life years and health status. *Oper Res.* 1980;28:206–224.
5. Drummond MF, O'Brien B, Stoddart GL, Torrance GW. *Methods for the Economic Evaluation of Health Care Programmes.* Oxford University Press; 2002.
6. Bleichrodt H, Quiggin J. Characterizing QALYs under a general rank dependent utility model. *J Risk Uncertainty.* 1997;15:151–165.
7. Bala MV, Wood LL, Zarkin GA, Norton EC, Gafni A, O'Brien BJ. Are health states "timeless"? The case of the standard gamble method. *J Clin Epidemiol.* 1999;52(11):1047–1053.
8. Franic DM, Pathak DS, Gafni A. Quality-adjusted life years was a poor predictor of women's willingness to pay in acute and chronic conditions: results of a survey. *J Clin Epidemiol.* 2005;58(3):291–303.
9. Lakdawalla D, Malani A, Reif J. The insurance value of medical innovation. *J Public Econ.* 2017;145:94–102.
10. Shafrin J, Schwartz TT, Okoro T, Romley JA. Patient versus physician valuation of durable survival gains: implications for value framework assessments. *Value Health.* 2017;20(2):217–223.
11. Lakdawalla DN, Romley JA, Sanchez Y, Maclean JR, Penrod JR, Philipson T. How cancer patients value hope and the implications for cost-effectiveness assessments of high-cost cancer therapies. *Health Aff (Millwood).* 2012;31(4):676–682.
12. Johannesson M. Theory and methods of economic evaluation of health care. *Dev Health Econ Public Policy.* 1996;4:1–245.
13. Bala MV, Wood LL, Zarkin GA, Norton EC, Gafni A, O'Brien B. Valuing outcomes in health care: a comparison of willingness to pay and quality-adjusted life-years. *J Clin Epidemiol.* 1998;51(8):667–676.
14. Torbica A, Tarricone R, Drummond M. Does the approach to economic evaluation in health care depend on culture, values, and institutional context? *Eur J Health Econ.* 2018;19(6):769–774.
15. Bleichrodt H, Herrero C, Pinto JL. A proposal to solve the comparability problem in cost-utility analysis. *J Health Econ.* 2002;21(3):397–403.
16. Feeny DH, Torrance GW. Incorporating utility-based quality-of-life assessment measures in clinical trials. Two examples. *Med Care.* 1989;27(3 suppl):S190–S204.
17. Lenert LA, Cher DJ, Goldstein MK, Bergen MR, Garber A. The effect of search procedures on utility elicitations. *Med Decis Making.* 1998;18(1):76–83.
18. Jansen SJT, Stiggelbout AM, Wakker PP, et al. Patients' utilities for cancer treatments: a study of the chained procedure for the standard gamble and time tradeoff. *Med Decis Making.* 1998;18(4):391–399.
19. Johannesson M. A note on the relationship between ex ante and expected willingness to pay for health care. *Soc Sci Med.* 1996;42(3):305–311.
20. Hammitt JK. QALYs versus WTP. *Risk Anal.* 2002;22(5):985–1001.
21. Dolan R, Edlin R. Is it really possible to build a bridge between cost-benefit analysis and cost-effectiveness analysis? *J Health Econ.* 2002;21(5):827–843.
22. Bleichrodt H, Quiggin J. Life-cycle preferences over consumption and health: when is cost-effectiveness analysis equivalent to cost-benefit analysis? *J Health Econ.* 1999;18(6):681–708.
23. Klose T. A utility-theoretic model for QALYs and willingness to pay. *Health Econ.* 2003;12(1):17–31.
24. Grossman M. On the concept of health capital and the demand for health. *J Political Economy.* 1972;80(2):223–255.
25. Mehrez A, Gafni A. Quality-adjusted life years, utility-theory, and healthy-years equivalents. *Med Decis Making.* 1989;9(2):142–149.
26. Mehrez A, Gafni A. The healthy-years equivalents: how to measure them using the standard gamble approach. *Med Decis Making.* 1991;11(2):140–146.
27. Gafni A, Birch S, Mehrez A. Economics, health and health economics: HYEs (healthy-years equivalent) versus QALYs (quality-adjusted live-year). *J Health Econ.* 1993;12(3):325–339.
28. Gafni A, Birch S. Preferences for outcomes in economic evaluation: an economic approach to addressing economic problems. *Soc Sci Med.* 1995;40(6):767–776.
29. Buckingham K. A note on HYE (healthy year equivalent). *J Health Econ.* 1993;12(3):301–309.
30. Culyer AJ, Wagstaff A. QALYs vs HYEs. *J Health Econ.* 1993;12(3):311–323.
31. Johannesson M, Pliskin JS, Weinstein MC. Are healthy-years equivalents an improvement over quality-adjusted life years? *Med Decis Making.* 1993;13(4):281–286.
32. Loomes G. The myth of the HYE. *J Health Econ.* 1995;14(1):1–7.
33. Bleichrodt H. QALYs and HYEs: under what conditions are they equivalent? *J Health Econ.* 1995;14(1):17–37.
34. Rittenhouse BE. Healthy years equivalents versus time trade-off. Ambiguity on certainty and uncertainty. *Int J Technol Assess Health Care.* 1997;13(1):35–48.
35. Mishan EJ. *Cost-Benefit Analysis.* New expanded ed. New York, NY: Praeger; 1976.
36. Prieto L, Sacristan J. Problems and solutions in calculating quality-adjusted life years (QALYs). *Health Qual Life Outcomes.* 2003;1:80.
37. Van Houtven G, Johnson FR, Kilambi V, Hauber AB. Eliciting benefit-risk preferences and probability-weighted utility using choice-format conjoint analysis. *Med Decis Making.* 2011;31(3):469–480.
38. Johnston RJ, Boyle KJ, Adamowicz W, et al. Contemporary guidance for stated preference studies. *J Assoc Environ Resour Econ.* 2017;4(2):319–405.
39. Lancsar E, Swait J. Reconceptualising the external validity of discrete choice experiments. *Pharmacoeconomics.* 2014;32(10):951–965.
40. Moore MJ, Viscusi WK. *Compensation Mechanisms for Job Risks: Wages, Workers' Compensation, and Product Liability.* Princeton, NJ: Princeton University Press; 1990.
41. Johnson FR, Hauber AB, Ozdemir S. Using conjoint analysis to estimate healthy-year equivalents for acute conditions: an application to vasomotor symptoms. *Value Health.* 2009;12(1):146–152.
42. Hauber AB, Itzler R, Johnson FR, et al. Healthy-days time equivalents for outcomes of acute rotavirus infections. *Vaccine.* 2011;29(45):8086–8093.
43. J.D. Lewis, M. Bewtra, F.I. Scott, et al. Anti-TNF drugs versus long-term steroid use for patients with inflammatory bowel diseases. Patient-Centered Outcomes Research Institute (PCORI). https://doi.org/10.25302/11.2018.CE.12114143.
44. Scott FI, Reed Johnson F, Bewtra M, et al. Improved quality of life with anti-TNF therapy compared to continued corticosteroid utilization in Crohn's disease. *Inflammatory Bowel Dis.* 2019;25(5):925–936.
45. Bridges JF, Hauber AB, Marshall D, et al. Conjoint analysis applications in health—a checklist: a report of the ISPOR good research practices for conjoint analysis task force. *Value Health.* 2011;14(4):403–413.
46. Gregor JC, McDonald JW, Klar N, et al. An evaluation of utility measurement in Crohn's disease. *Inflamm Bowel Dis.* 1997;3(4):265–276.
47. McFadden DL. Structural Discrete Probability Models Derived from Theories of Choice, in Structural Analysis of Discrete Data and Econometric Applications. In: Manski CF, McFadden DL, eds. Cambridge, MA: The MIT Press; 1981.
48. Patient Preference Information – Voluntary Submission, Review in Premarket Approval Applications, Humanitarian Device Exemption Applications, and De Novo Requests, and Inclusion in Decision Summaries and Device Labeling. U. S. Department of Health and Human Services, Food and Drug Administration. https://www.fda.gov/downloads/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/UCM517504.pdf.
49. Soekhai V, de Bekker-Grob EW, Ellis AR, Vass CM. Discrete choice experiments in health economics: past, present and future. *Pharmacoeconomics.* 2019;37(2):201–226.
50. Oppe M, Devlin NJ, van Hout B, Krabbe PF, de Charro F. A program of methodological research to arrive at the New International EQ-5D-5L valuation protocol. *Value Health.* 2014;17(4):445–453.