How adjusting elicited health utilities after the fact can adversely affect shared decision making [version 1; peer review: awaiting peer review]

Edouard Kujawski

Independent researcher, Berkeley, CA, 94708, USA

Abstract

Background: The elicitation of inconsistent health-state utility values (HSUVs) is a prevalent problem. There are two approaches to address this problem: (1) intervention during the elicitation process to ensure that patients estimate consistent HSUVs; (2) no intervention during the elicitation process and inconsistent HSUVs are adjusted after the fact. This paper studies three models recently proposed for adjusting inconsistent HSUVs and consistent HSUVs that some may consider unrealistic.

Analysis: The three models are analyzed using a sound theoretical framework: the mathematical equivalence of HSUVs elicited using the standard gamble and probabilities, the Fréchet bounds, and preference theory. It is proven that none of these models accounts for the Fréchet lower bound and health conditions that are preference substitutes.

Results: A clinical vignette proves these models may recommend treatments that result in premature death over treatments that cause acceptable adverse effects.

Conclusions: The three models are incorrect and may mislead patients and physicians to poor medical decisions. In the spirit of shared decision making, patients should be given the opportunity to reassess inconsistent HSUVs and confirm that the revised HSUVs reflect their preferences.

Keywords

Elicitation intervention, Fréchet bounds, Health-state utility values, Preference complements, Preference substitutes, Reasonableness test
Corresponding author: Edouard Kujawski (edkuj@comcast.net)

Author roles: Kujawski E: Conceptualization, Formal Analysis, Investigation, Methodology, Writing – Original Draft Preparation, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: The author(s) declared that no grants were involved in supporting this work.

Copyright: © 2022 Kujawski E. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Kujawski E. How adjusting elicited health utilities after the fact can adversely affect shared decision making [version 1; peer review: awaiting peer review] F1000Research 2022, 11:1533 https://doi.org/10.12688/f1000research.128862.1

First published: 19 Dec 2022, 11:1533 https://doi.org/10.12688/f1000research.128862.1
Introduction
The elicitation of inconsistent health-state utility values (HSUVs) is a prevalent problem. A rational person should not prefer a joint health state (JHS) to any of the constituent single health states (SHSs). This property is termed logical consistency. For instance, at the individual level, 41% of the HSUVs for the JHS (incontinence & impotence) elicited by Dale et al. violated this property. Given the mathematical equivalence of HSUVs elicited using the standard gamble (SG) and probabilities, logical consistency corresponds to the Fréchet upper bound (FUB). It is highly likely that the percentage of inconsistent HSUVs was higher than 41% because Dale et al. did not consider the Fréchet lower bound (FLB). The later has significant implications for the disutility of multiple coexisting morbidities: the joint disutility cannot exceed immediate death (ID) or the sum of the individual disutilities.

Two approaches to the problem of eliciting inconsistent HSUVs have been proposed:

1. Prevent inconsistent elicited HSUVs through interviewers intervening whenever necessary to ensure that patients estimate consistent HSUVs.
2. No intervention during the elicitation process. Inconsistent elicited HSUVs are adjusted after the fact.

The first approach requires trained and knowledgeable interviewers. The second approach requires realistic and mathematically valid models. However, the adjusted HSUVs may not accurately represent a patient’s preferences.

Triantaphyllou and Yanase (referred to as T-Y in this paper) proposed three models for adjusting inconsistent as well as consistent HSUVs which they say “may still not be realistic”:

- Model (i): “Readjusting the original health utility values via an error minimization approach based on the monotonicity property.”
- Model (ii): “Multiplicative functions for health states and a new model for adjusting the initial health-state utilities.”
- Model (iii): “A combined approach for adjusting the initial health-state utilities.”

This study has a sound theoretical framework: the mathematical equivalence of HSUVs elicited using SG and probabilities. It uses probability theory and preference theory to prove that the T-Y models are incorrect. A simple clinical vignette (Box 1) demonstrates that these models can be misleading. Therefore, they are inappropriate for shared decision making (SDM) where reliable HSUVs are critical for patients and physicians to decide upon a preferred treatment.

This paper proceeds as follows. The theoretical framework used for analyzing the T-Y models is presented. These models are analyzed, and it is demonstrated that they are inappropriate for life-critical SDM. Concluding remarks are presented.

Methods
Theoretical framework
Health-state utility values
Health states are identified by health conditions (HCs) (also termed “attributes” and “dimensions”) and severity levels. Each combination of levels of HCs represents a unique health state. HSUVs measure the strength of a person’s subjective preferences for health-related quality of life (HRQL). They are cardinal values specified on the (ID = 0.0, perfect health (PH) = 1.0) scale.

Estimating HSUVs is challenging. People are affected by the information which they receive, emotional factors, and elicitation methods. HSUVs elicited by different methods may not agree and can affect treatment choices. The SG has a theoretical foundation in von Neumann-Morgenstern expected utility theory, which establishes the validity of HSUVs elicited using the SG as a measure for HRQL. Using arbitrary scales for HSUVs can lead to serious errors.

Probabilities: Fréchet inequalities
Fréchet proved that the joint probability of two events is bounded by the marginal probabilities of each event regardless of the dependence between them. For two events A and B,
Few people realize they assign inconsistent values to joint probabilities. Osherson et al. state: "It is striking to observe, for example, how few people realize that it is inconsistent to attribute probabilities of 0.8 to each of two sentences and probability 0.5 to their conjunction." From (1), the conjunction of 0.8 and 0.8 cannot be less than 0.6: $P_{FLB}(0.8 \land 0.8) = 0.8 + 0.8 - 1.0$.

**Mathematical equivalence of HSUVs and probabilities**

In the SG, an individual is asked to make the hypothetical choice between living for $T$ years with health state $a_i$ (health condition $A$ with severity level $i$) and a gamble with a binary outcome (probability $p$ of living in PH for $T$ years or ID with probability $(1.0 - p)$). The probability $p$ is varied until the individual is indifferent between living $T$ years with $a_i$ and the gamble. The indifference probability $u_{a_i}(T)$ corresponds to the individual’s HSUV for $a_i$: $U_{a_i}(T) = u_{a_i}(T)$. Given the mathematical equivalence of HSUVs elicited using the SG and probabilities, probability theory can be applied to the problem of identifying inconsistent HSUVs.

**Consistent HSUVs: Fréchet inequalities**

Given that HSUVs are mathematically equivalent to probabilities, the Fréchet inequalities play an important role in identifying inconsistent HSUVs. $U(a_i, b_j)$ is bounded by the FUB and FLB on conjunction irrespective of preference interactions: $\max(0.0, U(a_i) + U(b_j) - 1.0) \leq U(a_i, b_j) \leq \min(U(a_i), U(b_j))$.

The FUB (2b) ensures logical consistency. The FLB (2c) has significant implications for the disutility of multiple coexisting morbidities. The joint disutility cannot exceed ID or the sum of the individual disutilities: $\overline{U}_{FLB}(a_i, b_j) = \min(1.0, U(a_i) + U(b_j) - 1.0)$.

**Preference interactions**

HCs can be mutually utility independent (MUI), preference complements (PCs), or preference substitutes (PSs). If a patient’s preference for condition $A$ is independent of the level of condition $B$ and vice versa, $A$ and $B$ are said to be MUI: $U^{MUI}(a_i, b_j) = U(a_i) \times U(b_j)$. If a patient believes that both $A$ and $B$ need to improve for their HRQL to improve, $A$ and $B$ are said to be PSs.
are said to be PCs. PC HSUVs are positively correlated: \( \text{UMUI}_{ai&b}^j < \text{C0/C1} \leq \text{FUB} \). If a person believes that only A or only B needs to improve for their HRQL to improve, A and B are PSs. PS HSUVs are negatively correlated: \( \text{FLB} \leq \text{UPS}_{ai&b}^j < \text{UMUI}_{ai&b}^j \):

### Quality-adjusted life-years

The linear quality-adjusted life-year (simply termed the QALY) is presently the principal model for medical decision making (MDM). The expected number of QALYs for living \( y_i \) years in a health state \( x_i \) which has a probability of occurrence \( p(x_i) \) is:

\[
\text{EQ}(p(x_i), x_i, y_i) = p(x_i) \times U(x_i) \times y_i + (1.0 - p(x_i)) \times y_i.
\]

### Analysis of T-Y models

In the following subsections, the T-Y models\(^4\) are analyzed using the above theoretical framework and data shown in Table 1.\(^1\) The analysis was done using Microsoft Excel 2016.

#### Model (i):

"Readjusting the original health utility values via an error minimization approach based on the monotonicity property."

The monotonicity property requires that a rational individual should not prefer a JHS to any of the constituent health states. Hence, Model (i) satisfies the FUB (2b). For consistency with probability theory, HSUVs elicited using the SG are also required to satisfy the FLB (2c). Model (i) does not address the FLB. For instance, it does not identify the HSUVs \( U_{ai} = 0.62, U_{bj} = 0.73, U_{ai&bj} = 0.15 \) as inconsistent: \( U_{ai&bj} < \text{FLB} (0.62 \times 0.73) = 0.35 (= 0.62 + 0.73 - 1.0) \).

#### Model (ii):

"Multiplicative functions for health states and a new model for adjusting the initial health-state utilities."

Model (ii) posits \( U_{ii}^{(ii)}(a_i \& b_j) = U(a_i) \times U(b_j) \). This assumes that health conditions \( a_i \) and \( b_j \) are MUI.

Keeney & Raiffa advocated assuming MUI with the significantly important qualifier\(^15\): \"the utility independence assumptions are appropriate in many realistic problems\". Moreover, their focus was principally on decision making outside of the medical domain. More recently, Howard and Abbas wrote\(^16\): \"We have several issues with this type of \’utility independence\’ reasoning … Enforcing these \’utility independence\’ assumptions result in functional forms that are simple, but quite frequently they will not represent the preference of the decision maker.\"

Experimental studies have concluded that the multiplicative model is not a suitable model for JHSUVs.\(^17\) MUI is a strong assumption that is usually inappropriate for HSUVs.\(^18\)

#### Model (iii):

"A combined approach for adjusting the initial health-state utilities"

Model (iii) posits that JHSs have a level of utility independence controlled by a parameter \( 0.0 \leq \gamma \leq 0.1 \). Thus, the JHSUVs lie between the MUI HSUV and the FBU (2b) and they do not account for HCs that are PSs. For instance, given \( U_{ai} = 0.62 \) and \( U_{bj} = 0.73 \), Model (iii) predicts \( 0.45 \leq U_{iii}^{(iii)}(a_i \& b_j) \leq 0.62 \). This is wrong: HCs can be PSs, in which case \( 0.35 \leq U_{PS}(a_i \& b_j) < 0.45 \).

---

### Table 1. Elicited and adjusted HSUVs using T-Y models.

<table>
<thead>
<tr>
<th>Methods</th>
<th>( U(a_i) )</th>
<th>( U(b_j) )</th>
<th>( U(a_i &amp; b_j) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>No intervention</td>
<td>0.62</td>
<td>0.73</td>
<td>0.15</td>
</tr>
<tr>
<td>Model (i)</td>
<td>0.62</td>
<td>0.73</td>
<td>0.15</td>
</tr>
<tr>
<td>Model (ii)</td>
<td>0.62</td>
<td>0.73</td>
<td>0.45</td>
</tr>
<tr>
<td>Model (iii)</td>
<td>0.62</td>
<td>0.73</td>
<td>( e \in [0.45, 0.62] )</td>
</tr>
</tbody>
</table>

\( e \in [., .] \) lies in interval \([., .] \)
T-Y recommend using Models (ii) and (iii) for JHSUVs that “would easily pass the previous monotonicity test but could still be considered as not realistic.” This can mislead clinicians to recommend and patients to choose unwanted treatments. Case in point, a patient who wants to avoid treatments with HUVs ≤0.45 chooses treatment T_X based on U^{ii}(a_i&b_j)/C_0/C_1 ≥ 0.45:

\[ U^{ii}(a_i&b_j)/C_0/C_1 = 0.38 \]

These HUVs satisfy the FUB (= 0.55) and FLB (= 0.17).

Results and discussion
The clinical vignette in Box 1 is analyzed assuming HUVs that are elicited with and without intervention.

HSUVs elicited with intervention
Interviewers intervene when necessary to ensure that patient HP assesses consistent HSUVs which truly represent their preferences. Elicited single HSUVs (SSHSUVs) are not always more correct than elicited JHSUVs. HP adjusts the JHSUV and SHSUvs as shown in Box 1: \( U(a_i) = 0.55, U(b_j) = 0.62, U(a_i&b_j) = 0.38 \). These HSUVs satisfy the FUB (= 0.55) and FLB (= 0.17).

HSUVs elicited without intervention
Interviewers do not intervene during the elicitation of HSUVs. The no-intervention elicited HSUVs shown in Table 1 violate the FLB (= 0.35). As discussed above, the T-Y models do not identify these HSUVs as inconsistent. T-Y recommend using Models (ii) and (iii) for JHSUVs that “could still be considered as not realistic.” These models predict the significantly different JHSUVs shown in Table 1.

Decision analysis
For illustration, we consider the clinical vignette and data shown in Box 1 and Table 2, respectively. Patient HP has a complicated decision to make: “to be or not to be” treated with T_X? The expected number of QALYs for each alternative and set of HSUVs is calculated using (4). Table 2 summarizes the results and recommendations. The HSUVs elicited with and without intervention provide contradictory recommendations:

- HSUVs elicited with intervention. The prediction is: 15.0 YLs, 7.8 QALYS. The recommendation is “Yes T_x”.

- HSUVs elicited without intervention and adjusted after the fact. Model (i) predicts 15 YLs and 2.25 QALYS. Models (ii) predicts 15 YLs and 6.75 QALYS. Based on the number of QALYS, Models (i), and (ii) recommend “No T_x”. Model (iii) recommends either “No T_x” or “Yes T_x” depending on the control parameter \( \gamma \).

Kujawski et al.\(^{10}\) proposed an intuitive reasonableness test that decision models used for SMD should pass to qualify as SDM tools: “Can a treatment that results in premature death trump a treatment that causes acceptable adverse effects?” A “Yes” answer may mislead clinicians into recommending and patients into choosing decisions with unintended consequences. As shown in Table 2, the three T-Y models fail this test.

Conclusions
The elicitation of reliable HSUVs is critical to ensure medical decisions that patients truly prefer. As shown in this paper, the three T-Y models\(^{4,5}\) do not accurately account for individual preferences and the mathematical equivalence of HSUVs with probabilities elicited using the SG. Given consistent elicited HSUVs, it is not the function of clinicians to judge whether these are realistic or unrealistic.

<table>
<thead>
<tr>
<th>Options</th>
<th>YLs</th>
<th>Elicitation</th>
<th>P (a_i &amp; b_j)</th>
<th>U (a_i &amp; b_j)</th>
<th>QALYs</th>
<th>REC</th>
<th>RT</th>
</tr>
</thead>
<tbody>
<tr>
<td>No T_x</td>
<td>7.00</td>
<td>Intervention</td>
<td>1.00</td>
<td>1.00</td>
<td>7.00</td>
<td>Yes</td>
<td>T_x</td>
</tr>
<tr>
<td>Yes T_x</td>
<td>15.0</td>
<td>Intervention</td>
<td>1.00</td>
<td>0.52</td>
<td>7.80</td>
<td>Yes</td>
<td>T_x</td>
</tr>
<tr>
<td>No intervention</td>
<td></td>
<td>Model (i)</td>
<td>1.00</td>
<td>0.15</td>
<td>2.25</td>
<td>No T_x</td>
<td>F</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Model (ii)</td>
<td>1.00</td>
<td>0.45</td>
<td>6.75</td>
<td>No T_x</td>
<td>F</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Model (iii)</td>
<td>1.00</td>
<td>∈(0.45, 0.62)</td>
<td>∈(6.75, 9.30)</td>
<td>?</td>
<td>F</td>
</tr>
</tbody>
</table>


Table 2. Impact of HSUV elicitation and T-Y models on treatment recommendation.
A clinical vignette proves that the three T-Y models may recommend treatments that result in premature death over treatments that cause acceptable adverse effects. This is a sure sign that these models are faulty and can be misleading. Well-trained interviewers are still essential to elicit reliable HSUVs. Practical tools are being developed to assist with the assessment of HSUVs, e.g., Gambler II. The uncertainties of elicited HSUVs and calculated QALYs need to be addressed for sound SDM. Assuming point estimates causes false confidence in the analysis results.

Data availability

Underlying data


Data are available under the terms of the Creative Commons Attribution 4.0 International license (CC-BY 4.0).

References

The benefits of publishing with F1000Research:

• Your article is published within days, with no editorial bias
• You can publish traditional articles, null/negative results, case reports, data notes and more
• The peer review process is transparent and collaborative
• Your article is indexed in PubMed after passing peer review
• Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com