Estimating Preference Scores in Conventional and Home Nocturnal Hemodialysis Patients

Philip A. McFarlane,* Andreas Pierratos,† Ahmed M. Bayoumi,‡ and Donald A. Redelmeier§

*Home Dialysis and †Department of Medicine, Inner City Health Research Unit, St. Michael’s Hospital, ‡Home Dialysis, Humber River Regional Hospital, and §Department of Clinical Epidemiology, Canada Research Chair in Medical Decision Sciences, Sunnybrook and Women’s College Health Sciences Centre, Toronto, Ontario, Canada

Previous studies have reported higher quality of life in patients who receive home nocturnal hemodialysis (HNHD) than conventional in-center hemodialysis (IHD). The optimal method for eliciting preferences from dialysis patients remains undefined, and there may be unique methodologic concerns in this population. Patients’ preferences for IHD (n = 20) and HNHD (n = 24) were studied using the standard gamble (SG), time trade-off (TTO), and modified willingness to pay (WTP) methods. This report describes experience with operationalizing these three techniques in this population. A higher preference for HNHD was found with all measures, with significant differences observed with the SG (HNHD: median 0.79 [interquartile range (IQR) 0.67 to 0.95]; IHD: median 0.60 [IQR 0.20 to 0.82]; P = 0.031) and WTP (HNHD: median 0.50 [IQR 0.40 to 0.68]; IHD: median 0.20 [IQR 0.20 to 0.38]; P < 0.001). SG and TTO scores were moderately correlated but not with WTP. In addition, qualitative issues arose during TTO and WTP interviews that seemed to influence the interpretation of these preference scores. In the TTO, time willing to trade became oriented toward the next pivotal life event, with a failure of the requirement for a constant proportional time trade-off. WTP preferences were oriented toward the smallest survival stipend. These issues represent range restriction biases. No significant issues arose during the SG interviews. HNHD patients expressed a greater preference for current health than IHD patients. The operational performance of SG was good in this study, whereas biases and methodologic concerns were identified with the TTO and WTP in this population.


Measuring quality of life is an attractive approach to gauging an intervention’s effectiveness. Quality-of-life measurements can be a study’s primary focus or can be combined with costing data as part of a cost-effectiveness analysis. The quality of life of dialysis patients is among the worst reported for chronic medical conditions (1,2). Studies have attempted to quantify quality of life in dialysis patients through disease-specific or generic instruments (3–11). Tools such as the kidney disease quality of life (KDQOL) generate a numeric rating of quality of life in many domains. Preference-based techniques can be used to estimate utility scores (2,4,12,13), expressing global quality of life as a single number on a scale ranging from 0 (a health state equivalent to death) to 1 (equivalent to the best possible health (14–20). They have several advantages over rating systems, allowing comparisons of quality of life across disease states, and can be used to generate the popular quality-adjusted life year (QALY) metric that is used in cost-utility studies. Utility scores can be generated by a variety of preference-based methods and are one of the preferred outcome measures for economic analyses (21,22).

Two such methods are the time trade-off (TTO) and standard gamble (SG). The TTO method involves determining the maximum proportion of remaining lifespan that a patient would be willing to trade for perfect health. The SG method involves determining the maximum chance of immediate death that a patient would be willing to gamble against for the chance of gaining perfect health. Patient preferences can also be elicited through willingness to pay (WTP) valuations, particularly for application in cost–benefit analyses (23–26).

The preferred technique for eliciting preferences remains controversial, because preference scores can vary considerably (27). Such variation may be due to the heterogeneity of personal preferences. However, previous studies suggested three methodologic concerns that can also influence preference scores. First, preference assessment studies exhibit method effects: The technique used seems systematically to bias preference scores (19,28–30). Second, preference assessment studies exhibit respondent effects: Interviews of patients with the medical condition, health professionals, and lay public often yield different results (31–33). Finally, the validity of preference measures rest on axioms that can be violated (34,35).

Our group performed a cost-utility study that examined two forms of hemodialysis (2). The first was conventional in-center hemodialysis (IHD), whereby patients received dialysis treatments at the hospital three times per week for 4 h. The second was home nocturnal hemodialysis (HNHD), whereby patients...
performed treatments at home five or six times a week, connecting themselves at night to the dialysis machine and undergoing dialysis during sleep (36). In our primary report, we identified the cost-utility of HNHD relative to IHD, with the utility score based on the SG technique as the measure of dialysis effectiveness. However, we also elicited patient preferences for health states by the TTO and WTP methods, with an à priori intention to examine the performance of these measures in this population.

In this report, we focus on two areas that were not discussed in our previous publication. First, we describe the degree of correlation among these three techniques. In theory, these scores should be highly correlated, but in practice, this is often not the case (28). Second, we report an informal qualitative description of our experiences with eliciting preferences, including potential operational problems when these techniques are applied in hemodialysis patients. Because our sample size is relatively small, we cannot offer a rigorous generalizable comparison of these three techniques; however, our observations suggest areas for future research and point out potential pitfalls that future investigators will need to take into account. This information is particularly significant given the increasing popularity of many forms of intensive dialysis and the need to estimate the impact of these modalities on quality of life.

Materials and Methods

Patient preferences were elicited through face-to-face interviews that were conducted in a standardized manner by a single interviewer, with interviewer instructions written on a prompt card. Sessions were scheduled during a routine appointment at the home dialysis clinic for the HNHD group and at the dialysis unit before treatment for the IHD group. Interviews were conducted in well-lit and quiet semiprivate offices or examination areas. Criteria were set prospectively for delaying an interview; specifically, hospital admission within 1 mo, current treatment with antibiotics, initiation or dosage titration of a psychotropic drug within 1 mo, invasive procedure within 1 wk or, significant life event within 2 mo (death or sickness of a friend or family member, change in employment status, divorce or separation from significant other, or change of address). The interviews for each utility measure were performed sequentially, starting with TTO, then SG, then WTP. We did not vary the order of testing and did not attempt to identify sequencing biases.

Each elicitation task was framed as a dichotomous choice, in which participants were asked whether they preferred either the certainty of their current health or an alternative in which they could attain best possible health by accepting a trade-off. The trade-offs were risk for death, decreased life expectancy, and monetary payments for the SG, TTO, and WTP methods, respectively. The choice was from the ex-post user-based perspective. We assigned discrete values to these trade-off parameters: Probability of death for the SG, time traded for the TTO, and money lost for the WTP. For each method, the initial value was randomly chosen from all possible values. We changed the value of the trade-off parameter iteratively using a bisectional approach (in which the utilities presented were the midpoint of the possible lowest and highest values on the basis of the previous responses and the bounds of the preference scale) until we found the maximum value that each respondent would accept for the trade-off parameter (e.g., maximum probability of immediate death in the SG technique) (16,28). We calculated preference scores on the basis of this indifference point. We did not elicit preferences between IHD and HNHD, because not all patients had been on both modalities; we also did not attempt to compare these with other dialysis modalities (e.g., peritoneal dialysis) or renal transplantation.

For the SG, participants were given a choice between remaining in their current state of health or accepting a hypothetical treatment. If successful, then the hypothetical treatment would cure the renal disease, resolve their medical issues, and return the person to the best possible state of health. If unsuccessful, then the hypothetical treatment would lead to immediate death. Interviewers used a computer program to represent visually the gamble probabilities in the form of a circle, transected into two sectors, one representing the probability of immediate death and the other representing the probability of a successful cure (http://individual.utoronto.ca/bayoumi/prospect/excel.htm). Patients were allowed to choose not to gamble.

For the TTO technique, participants were given a choice between remaining in their current state of health or accepting a hypothetical treatment that would cure the person’s renal disease, resolve all medical issues, and restore the best possible state of health. However, the hypothetical treatment would reduce their remaining lifespan. The same computer-based utility elicitation program provided assistance, visually representing the time horizon as a horizontal bar whose length represented the patient’s remaining life span, bisected such that the bar consisted of two colored areas, one representing the proportion of the lifespan that will be spent in perfect health and the remaining area representing the proportion that would be lost to achieve perfect health. The estimated remaining lifespan was drawn from American life expectancy tables and adjusted for age, race and gender but not presence of renal or other diseases. Patients were allowed to choose not to trade off time.

Patient preferences were also elicited by two variations of the WTP method. The first approach involved an open-ended determination of the amount that a patient was willing to pay to achieve best possible health. In this case, the scale was bounded by zero and by the total of all of the patient’s available financial resources as reported by the respondent. For reasons discussed next, a second evaluation of WTP was also used. In the second approach, patients were assigned an after-tax income of $50,000 per year, with no possibility of acquiring new funds or of drawing on existing resources (making the scale bounded by $0 and $50,000). The payment for perfect health would be made annually, thereby reducing their annual income for the remainder of their lives. In both cases, we generated a WTP score equivalent to 1 minus the maximum acceptable payment divided by the scale maximum (1 − u, where u is maximum willingness to pay/scale maximum). Like the other measures, this calculation produces a WTP score that ranges from 0 to 1, with higher scores suggesting a better quality of life.

Statistical Analyses

We used nonparametric methods to evaluate preference scores. For continuous variables, between-group differences were tested with the Mann-Whitney U test. For categorical variables, between-group differences were tested using the χ2 test or Fisher exact test as necessary. Correlation between preference scores and demographic variables was determined using Kendall’s τ-b statistic and among the three forms of preference scores by calculating the intraclass correlation coefficient using the Cronbach α method for single measures with a two-way mixed-effect model, testing against absolute agreement with a moderate correlation threshold of 0.3 and a highly correlated threshold of 0.6. A two-tailed significance level of 0.05 was used for all tests. SPSS (version 11.0; SPSS, Chicago, IL) for Macintosh was used for the statistical computations.
Results
We interviewed 20 IHD and 24 HNHD patients (Table 1). Preference scores were higher in the HNHD group by all techniques (Table 2), although statistical significance was reached only for the SG and WTP techniques. Differences that were related to form of dialysis persisted after adjustment for demographic and comorbid factors. The TTO and SG scores were moderately correlated with each other, whereas the WTP scores did not significantly correlate with either of the other measures (Table 3, Figure 1). The dispersion of scores within groups by the three techniques was broad (Figure 2).

TTO Observations
Eleven of the 43 patients (six IHD, five HNHD; NS) expressed spontaneously that they wanted to live only until their next pivotal life event, were unwilling to trade any time before that event, and were willing to sacrifice all remaining time beyond that point. In one example from the interview logs, a respondent stated, “I only need to live until my nephew’s wedding next summer; after that, I don’t care.” Those who responded in this manner were not confronted about their response or reinterviewed later to determine whether their desired lifespan had decreased as predicted. TTO utility scores were lower in those who responded in this manner (0.18 ± 0.15 versus 0.69 ± 0.22; P < 0.001); however, no statistically significant demographic differences were apparent, although the sample size is small. Two people chose not to trade. One person had to reschedule the interview because the TTO questions were found to be upsetting, but the study was eventually completed. No other major issues arose during the TTO interviews.

WTP Observations
In the WTP analysis, seven respondents (five IHD, two HNHD) were interviewed using the first methodologic variation (scale upper bounded by all of the patient’s available financial resources). In each case, respondents answered similarly, expressing a willingness to trade everything with the exception of a small survival stipend. The protocol was then modified to the second methodologic variation (scale upper bounded by $50,000/yr). The original seven were reinterviewed, and this method was used for the remainder of the study group; these scores are reported here. Approximately one fourth of the participants spontaneously commented that the $50,000/yr figure was generous, whereas another one fourth observed that this amount seemed inadequate. All participants were willing to trade money for health.

SG Observations
No significant difficulties arose during the SG interviews. Two patients required a complete repeat of the instructions but were able to complete the interview without further difficulties. All patients were willing to gamble.

Overall Observations
No evidence of editing bias (whereby respondents have revised the provided scenario information) was observed, and the tasks did not seem to be too cognitively demanding or incomprehensible in our population. Subjectively, there was no evidence of a social desirability bias (whereby respondents provide what they perceive to be the desired response rather than their true valuation). The average time to complete each set of elicitations was 4 min (not including time for instructions). When an individual rated his or her quality of life very highly (>0.90 on any metric), we challenged the individual about the willingness to trade. In all cases, the individuals agreed that they were willing to trade or gamble in principle but rated their current quality of life highly enough that they were willing to trade or gamble only small amounts or nothing.

Discussion
A variety of methods are available to elicit patient preference for health states (34,35). To be consistent with theoretical underpinnings, several requirements must be met (14,37–40). In our study, three methods of determining preference were applied, and observations were made with two of these methods that raise cautions when they are used in hemodialysis patients. Although it is possible that these issues are present in other groups, we restrict our conclusions to the population that we studied. We have attempted to provide some statistical rigor;

Table 1. Demographic profile of hospital- and home-treated patientsa

<table>
<thead>
<tr>
<th>Parameter</th>
<th>IHD</th>
<th>HNHD</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>19</td>
<td>24</td>
<td></td>
</tr>
<tr>
<td>Gender (% male)</td>
<td>68</td>
<td>75</td>
<td>0.63</td>
</tr>
<tr>
<td>% with diabetes</td>
<td>11</td>
<td>8</td>
<td>0.81</td>
</tr>
<tr>
<td>% with coronary artery disease</td>
<td>11</td>
<td>21</td>
<td>0.35</td>
</tr>
<tr>
<td>% with peripheral vascular disease</td>
<td>5</td>
<td>13</td>
<td>0.42</td>
</tr>
<tr>
<td>% with congestive heart failure</td>
<td>5</td>
<td>21</td>
<td>0.14</td>
</tr>
<tr>
<td>% with postsecondary education</td>
<td>69</td>
<td>60</td>
<td>0.71</td>
</tr>
<tr>
<td>% married</td>
<td>54</td>
<td>94</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Age (yr; mean [SD])</td>
<td>50.1 ± 9.3</td>
<td>47.2 ± 7.7</td>
<td>0.26</td>
</tr>
<tr>
<td>Duration of ESRD (yr; mean [SD])</td>
<td>7.1 ± 6.0</td>
<td>9.4 ± 7.1</td>
<td>0.27</td>
</tr>
</tbody>
</table>

aHNHD, home nocturnal hemodialysis; IHD, in-center conventional hemodialysis.
however, the small sample size lends itself better to a subjective
description of the biases observed.

Our most serious caution arose with the TTO method, in
which some participants traded all remaining lifetime aside
from the amount necessary to reach their next pivotal life event,
an interval that was usually arbitrary and outside the control of
the patient yet had an undue effect on the TTO score. This
pattern could seriously distort the interpretation of TTO scores
in such respondents. For example, if a participant wanted to
live only until a pivotal event and that event date was acceler-
ated by 6 mo, then the amount of time that the patient would be
willing to trade would (presumably) be increased by 6 mo.
Although the TTO score would decrease, it seems unlikely that
there was a change in the patient’s underlying quality of life or
of their preference for health states. Although the utility func-
tion has been described as being concave over time in most
people (30,41,42), those who responded in this manner would
have a step function for utility over time, which would repre-
sent both a range restriction bias and an extreme example of
nonconstant proportional TTO. Although other authors have
suggested ways to correct TTO scores by adjusting for time
preference bias (42), these methods would not apply in our
extreme cases. When the time willing to trade is fixed by a
pivotal life event, differences in preference scores between in-
dividuals are primarily determined by the expected remaining
lifespan of 10 and 20 yr, respectively, both with a pivotal life
event 1 yr away. Although the first patient would seem to have
a higher TTO score (0.10 versus 0.05), it is not clear that the first
patient is actually expressing a higher preference for dialysis.
Again, such patterns represent violations of the requirement for
constant proportional TTO (43–45) and are a form of range
restriction bias.

Although possible, we find it unlikely that our operational-
ization of the TTO method contributed to these biases. Our TTO
technique was similar to that used in previous ESRD studies,
and our IHD population had similar utility scores to those seen
in previous dialysis studies. We did not structure the inter-
views in a way that would have encouraged spontaneous ex-
pressions such as we observed. It is possible that these obser-
vations applied only to our unique population or were related
to our operational methods; however, it is also possible that
participants in previous trials harbored similar thoughts.

Difficulties also arose in our WTP analysis. Participants
seemed willing to trade all but a small stipend necessary for
survival. This calls into question whether a similar range re-
striction bias was present in the WTP interviews, with values
being dictated by a predetermined minimal necessary income
level rather than allowing for a clear expression of a partici-
 pant’s preferences. This observation could lead to difficulties in
interpreting the WTP score. For example, consider two such
people studied with our first methodologic variation (with the

<table>
<thead>
<tr>
<th>Preference Measure</th>
<th>IHD</th>
<th>HNHD</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>SG median</td>
<td>0.79</td>
<td>0.60</td>
<td>0.031</td>
</tr>
<tr>
<td>IQR</td>
<td>0.67 to 0.95</td>
<td>0.20 to 0.82</td>
<td></td>
</tr>
<tr>
<td>TTO median</td>
<td>0.62</td>
<td>0.58</td>
<td>0.21</td>
</tr>
<tr>
<td>IQR</td>
<td>0.38 to 0.91</td>
<td>0.27 to 0.67</td>
<td></td>
</tr>
<tr>
<td>WTP median</td>
<td>0.50</td>
<td>0.20</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>IQR</td>
<td>0.40 to 0.68</td>
<td>0.20 to 0.38</td>
<td></td>
</tr>
</tbody>
</table>

*IQR, interquartile range; SG, standard gamble; TTO, time trade-off; WTP, willingness to pay.

<table>
<thead>
<tr>
<th>Preference Measure</th>
<th>Intraclass Correlation Coefficienta</th>
<th>Moderately Correlatedb</th>
<th>Highly Correlatedc</th>
</tr>
</thead>
<tbody>
<tr>
<td>SG with TTO</td>
<td>0.65 (0.41 to 0.80)</td>
<td>0.001</td>
<td>0.70</td>
</tr>
<tr>
<td>with WTP</td>
<td>0.25 (−0.03 to 0.51)</td>
<td>0.64</td>
<td>1.0</td>
</tr>
<tr>
<td>TTO with WTP</td>
<td>0.16 (−0.12 to 0.43)</td>
<td>0.84</td>
<td>1.0</td>
</tr>
</tbody>
</table>

aWith 95% confidence interval.
bTested against correlation threshold of 0.3.
cTested against correlation threshold of 0.6.
patient’s total available financial resources as the upper bound of the scale), whereby one individual is much wealthier, but both are willing to trade all but the same survival stipend. It is not clear that the wealthier individual (whose WTP score would be lower) is actually expressing a different preference for the dialysis state than the less wealthy patient, despite the difference in scores. In the second WTP method (whereby the scale was upper bounded by $50,000/yr), there was significant disparity in patients’ response to the availability of $50,000/yr of income, perhaps related to the declining marginal utility of a dollar to wealthy rather than poor individuals (26). These issues call into question the ability of WTP methods as we applied them to differentiate patient preferences for health when quality of life is extremely poor. However, we believe that the WTP approach has promise in dialysis populations, because our participants were able to perform the required cognitive tasks and seemed willing to trade money for health. Future research should examine which WTP method performs best in this population. For example, WTP valuations that involve comparisons of one dialysis modality with another, rather than with a hypothetical treatment that is capable of normalizing health, may produce better results. Also, structured analyses that compare open- versus closed-ended scenarios, as well as different approaches to the bargaining method, should be examined. These approaches may better represent patient preference for health states.

Unlike the TTO and WTP methods, no significant issues were identified during the SG analysis. Although it has been suggested that the abstract thinking that is required for the SG method may make it difficult to use (38), we did not observe such difficulties. It is possible that difficulties with the SG may arise in other dialysis populations, or perhaps people who perform dialysis are particularly well suited to consider hypothetical life or death decisions and numeric expressions of probability.

Although it seems that individuals with high preference scores were willing to trade or gamble, it is possible that in some cases there was a true unwillingness to trade or gamble that would become evident on more detailed questioning. Previous studies in dialysis (46) and in nonrenal populations (19,28–30) suggested that preference scores are often highest with SG, then TTO, and lowest with WTP. Factors such as utility curvature, probability weighting, loss aversion, and scale compatibility lead to these systematic differences among methods (30). We saw a similar pattern in our results, suggesting that method effects were active in our population.

Figure 1. Scatter plots comparing the preference scores from a cohort of hemodialysis patients. (A) Standard gamble (SG) versus time trade-off (TTO). (B) Willingness to pay (WTP) versus SG. (C) WTP versus TTO.

Figure 2. Distribution of preference scores by type of dialysis and preference measurement technique. The top, bottom, and line through the middle of the box correspond to the 75th, 25th, and 50th (median) percentiles, respectively. The whiskers on top and below the box extend to the 90th and 10th percentiles, respectively. The black square within the box corresponds to the mean.
Our observations are based on spontaneous comments made by participants. No attempt was made to probe whether these patterns of reasoning were present in others who did not volunteer similar comments, so our observations may underestimate their prevalence. These potential biases may be present in other populations, in particular other health conditions in which quality of life is chronically low. Regardless of the theoretical issues that arise from these observations, the desire to live only to the next pivotal life event and the willingness to trade all but a minimal survival stipend are clear reminders of how poor the quality of life may be for many people who are on dialysis.

A final observation is the impressive variability in utility scores between individuals who perform the same method of dialysis, sometimes resulting in a range that covers the entire utility scale. Investigators who perform studies of quality of life in dialysis patients should be prepared to see similar variability in their results, which may have an impact on patient selection and study sample size.

Conclusion
We measured patient preferences for health states by three methods, all of which suggested a higher quality of life in patients who performed HNHD, with statistically significant differences seen for the SG and WTP techniques. Each preference elicitation technique has advantages and disadvantages. The TTO method may be vulnerable to biases as we have described. The SG method worked well in our study, but the cognitive demands of this technique may be problematic in other dialysis populations. The WTP techniques that we used did not perform well but remain appealing options that will require further refinement. Although the measurement of patient preference for health states remains an appropriate method of estimating quality of life, further research is required to define the best method of elicitting such preferences in people with ESRD.

Acknowledgments
The Canadian Society of Nephrology and the Kidney Foundation of Canada provided research support for the conduction of this study and for P.A.M. Funding for HNHD pilot programs provided by the Ontario Ministry of Health. The funding agreement ensured the authors’ independence in designing the study, interpreting the data, writing, and publishing the report.

We acknowledge the following institutions and people for participation: St. Michael’s Hospital: Jill Campbell, Mark Goldstein, and Mark Robson; Humber River Regional Hospital: Colleen Jones and Michaelene Ouwendyk.

Disclosures
None.

References
19. Blumenschein K, Johannesson M: Relationship between quality of life instruments, health state utilities, and will-