The relationship of medical, demographic and psychosocial factors to direct and indirect health utility instruments in rheumatoid arthritis

A. G. Witney1,2, G. J. Treharne2,1, M. Tayakoli3, A. C. Lyons4, K. Vincent5, D. L. Scott5 and G. D. Kitas1,2

Objectives. Cost-effectiveness analysis (CEA) is essential for the comparison of treatments for rheumatoid arthritis (RA). CEA centres on accurate measurement of health utility (HU) preferences. Direct measures of HU in RA patients demonstrate weaker correlations with health status (functional disability and pain) than indirect measures. We examined whether demographic and psychosocial factors relate to HU in RA patients.

Methods. HU was measured for 142 RA patients (76% women; mean age 58.75 yr) directly through standard gamble (SG) and time trade-off (TTO), and indirectly on the EuroQol (EQ-5D). Current pain (100 mm visual analogue scale) and recent functional disability (Health Assessment Questionnaire; HAQ) were assessed. A subsample of 48 provided demographic and psychosocial information (education, employment, marital/family status, knowledge about RA, medication beliefs, desirable responding, social support, optimism, and the Hospital Anxiety and Depression Scale; HADS).

Results. Direct HU had higher means (SG = 0.88, TTO = 0.86) than indirect HU (EQ-5D = 0.52). HAQ functional disability correlated with SG (r = −0.28), TTO (r = −0.31) and EQ-5D (r = −0.67). Current pain correlated with TTO (r = −0.19) and EQ-5D (r = −0.36). HADS depression correlated with TTO (r = −0.35) and EQ-5D (r = −0.64); HADS anxiety also correlated with EQ-5D (r = −0.46).

Conclusions. Demographic and psychosocial factors cannot completely explain either the significant differences between direct and indirect HUs in RA patients or the moderate correlations of direct HUs with health status. Characteristics of the SG and TTO may make them inappropriate for HU assessment and CEA among RA patients.

Key words: Rheumatoid arthritis, Health utility, Standard gamble, Time trade-off, Disability evaluation.
public healthcare systems, such as the British National Health Service (NHS), the assessment of the utility of health states should be determined by the public (the tax payers) rather than individual patients, making indirect methods preferable.

Recent studies [11, 13] suggest that there are considerable variations in HUs across different indirect measures, but have not compared these with direct HU instruments. It is important to understand why the same HUs are not elicited by different measures.

One of the problems with direct measures of HU is the assumption that decision-makers are rational individuals with no reference to demographic or psychosocial context. Furthermore, while questionnaires differ clearly in the dimensions used to construct a health state [9–11], it is not clear what attributes patients include in the construct of their health state in SG and TTO [16, 17].

This study is the first to examine in detail the health state construct underlying the main HU measures among people with RA. If demographic and psychosocial factors relate to patients’ willingness to gamble or trade on their current health state, then this may explain the lower correlation than would be expected between HU and other measures of health status (i.e. pain and functional disability in RA). Previous research has suggested that higher socio-economic status or education and being married or living with family lead to risk aversion [18, 19]. Any effect of education may be due to lack of understanding of the relatively complex probabilistic procedures [14]. Furthermore, the gamble or trade-off offered may be unconvincing for the patients with a chronic but non-terminal condition, such as RA, due to the human principle of risk aversion [20]. Over time, RA patients may become more tolerant of the disease and thus adapt to living in health states that would elicit a low HU among the general population [14]. Knowledge about RA may also be a factor in risk avoidance [21], as may patients’ medication beliefs [22], response style [23], social situation [9–11], optimism [16] or depression/anxiety [11, 12].

We hypothesized that demographic or psychosocial factors relate to the health state construct elicited by direct measures (SG, TTO) of HU among RA patients and carried out an explorative correlational study to provide an initial test of this. This provides an essential first step for understanding how to appropriately compute HU among people with RA, and inform valid CEA of treatment interventions.

**Methods**

**Participants**

One hundred and fifty patients classified with RA [24] were consecutively recruited from rheumatology out-patient clinics of the Dudley Group of Hospitals (West Midlands, UK) and King’s College Hospital (London, UK), where local research ethics committee approval was granted. Participants were given a written description of the project and signed consent forms. Two female researchers performed the interviews for the SG, TTO and EQ-5D, one at the Dudley Group of Hospitals (80 patients) and one at King’s College Hospital (70 patients). Order of administration of the SG, TTO and EQ-5D was systematically counterbalanced (i.e. varied equally) across the sample. Eight of the 150 participants (5%) were unable to understand the procedure; results refer to the remaining 142 participants, of whom 108 (76%) were female and the mean age was 58.75 yr (S.D. 13.25; range 16–87). A subsample of 50 patients from Dudley participated in a more detailed project that involved collection of demographic and psychosocial data. The questionnaires were self-completed at home by the participants and returned by post, thus reducing possible fatigue effects in the administration of further measures in the main interview containing the HU assessments. Two participants in this subsample (4%) were unable to understand the procedure (not a significant difference from the remaining sample; Fisher’s exact $P = 0.28$); in this part of the study all results refer to the remaining sample of 48 participants. The participants in the more detailed study did not differ in age (Table 1; $P = 0.26$) or sex (Table 2; $P = 0.06$) from the remaining sample.

**Measures employed in all participants**

*The standard gamble (SG)*. SG offers participants a gamble between staying in their current state of health for the rest of their lives or taking a hypothetical medication that has a chance of causing instant and permanent perfect health, but with a risk of death [4–6]. The risk of death ($q = 1−p$, where $p$ is the probability of successful treatment) was initially presented as 5% and gradually increased in increments of 5% and then 10% (with the opportunity for smaller increments) until participants could not decide whether or not they would take the hypothetical medication.

**Table 1. Descriptive statistics of medical, demographic and psychosocial factors and correlations with health utility scores**

<table>
<thead>
<tr>
<th>Variable (measure)</th>
<th>Descriptive statistics</th>
<th>Correlation ($r$) with</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>s.d.</td>
</tr>
<tr>
<td>Functional disability (HAQ)</td>
<td>1.39</td>
<td>0.88</td>
</tr>
<tr>
<td>Pain (100 mm VAS)</td>
<td>35.70</td>
<td>22.85</td>
</tr>
<tr>
<td>Inflammation (CRP)</td>
<td>11.93</td>
<td>13.63</td>
</tr>
<tr>
<td>RA duration (years)</td>
<td>8.35</td>
<td>10.70</td>
</tr>
<tr>
<td>Age (years)</td>
<td>56.98</td>
<td>14.02</td>
</tr>
<tr>
<td>Education (years)</td>
<td>15.65</td>
<td>1.56</td>
</tr>
<tr>
<td>Number of children (at home)</td>
<td>1.67</td>
<td>1.77</td>
</tr>
<tr>
<td>Knowledge of RA (PKQ)</td>
<td>17.67</td>
<td>4.21</td>
</tr>
<tr>
<td>Perceptions of necessity (BMQ)</td>
<td>11.26</td>
<td>3.05</td>
</tr>
<tr>
<td>Perceptions of concern (BMQ)</td>
<td>14.81</td>
<td>3.84</td>
</tr>
<tr>
<td>Perceptions of overuse (BMQ)</td>
<td>12.83</td>
<td>2.16</td>
</tr>
<tr>
<td>Perceptions of harm (BMQ)</td>
<td>14.68</td>
<td>1.91</td>
</tr>
<tr>
<td>Desirable responding (BDIR)</td>
<td>6.56</td>
<td>4.43</td>
</tr>
<tr>
<td>General social support (SSS)</td>
<td>3.08</td>
<td>0.72</td>
</tr>
<tr>
<td>Medical social support (SSS)</td>
<td>3.35</td>
<td>0.61</td>
</tr>
<tr>
<td>Optimism (LOT)</td>
<td>18.35</td>
<td>5.12</td>
</tr>
<tr>
<td>Depression (HADS)</td>
<td>6.04</td>
<td>4.03</td>
</tr>
<tr>
<td>Anxiety (HADS)</td>
<td>8.02</td>
<td>4.88</td>
</tr>
</tbody>
</table>

* $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$. 
Time trade-off (TTO). TTO involves participants imagining two hypothetical courses of life that lie ahead: remaining in their current state of health for the maximum number of years (from a set allocation), or a shorter life spent in perfect health [4–6]. Different hypothetical lives (of t yr) were offered dependent on the age of the participant. Participants aged <40 were asked to trade off from 40 yr of life; 40–50-yr-olds were presented with 30 yr; 50–70-yr-olds were presented with 20 yr and participants aged >70 yr were presented with 10 yr. Options were presented verbally on cards that allowed participants to indicate the longest amount of time (if any) they would be prepared to trade off \((t − x)\) to have \(x\) yr in perfect health \([91\text{ participants (51%) refused to trade any time. The HU is the proportion of years in perfect health \((x/t)\). This provision of props and interview administration during SG and TTO assessments aids participant comprehension [25].\]

The EuroQol (EQ-5D). EQ-5D [12] questions self-assessed problems across five items on mobility, self-care, usual activities, pain/discomfort and depression/anxiety. Each item has three levels of severity: ‘no problems’, ‘some problems’ and ‘severe problems’. This provides 243 possible health states. HU is computed by applying a weighting system derived from a UK population-based survey, where participants used the TTO to rate HU for a proportion of EQ-5D health states [12].

Functional disability. The Anglicized version of the 40-item Stanford Health Assessment Questionnaire (HAQ) [26] was used to measure functional disability. Participants rate their ability (over the past week) to carry out 20 activities within eight aspects of daily living on a four-point scale. The HAQ is a commonly employed, reliable and valid measure of functional disability [27–29].

Pain. A 100 mm visual analogue scale (VAS) was used to measure current severity of pain (i.e. on the day) [30].

### Table 2. Comparison of health utility scores by demographic factors

<table>
<thead>
<tr>
<th>Variable groups</th>
<th>Standard gamble (SG)</th>
<th>Time trade-off (TTO)</th>
<th>EuroQol (EQ-5D)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>s.d.</td>
<td>Mean</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male (n = 15)</td>
<td>0.92</td>
<td>0.07</td>
<td>0.90</td>
</tr>
<tr>
<td>Female (n = 33)</td>
<td>0.87</td>
<td>0.14</td>
<td>0.85</td>
</tr>
<tr>
<td>Education</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Primary only (n = 19)</td>
<td>0.90</td>
<td>0.11</td>
<td>0.89</td>
</tr>
<tr>
<td>Secondary or above (n = 29)</td>
<td>0.87</td>
<td>0.14</td>
<td>0.85</td>
</tr>
<tr>
<td>Employment</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>None paid (n = 32)</td>
<td>0.86</td>
<td>0.14</td>
<td>0.83</td>
</tr>
<tr>
<td>Paid (full or part time) (n = 16)</td>
<td>0.93</td>
<td>0.07</td>
<td>0.93</td>
</tr>
<tr>
<td>Marital status</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not married (n = 10)</td>
<td>0.91</td>
<td>0.09</td>
<td>0.89</td>
</tr>
<tr>
<td>Married (n = 38)</td>
<td>0.88</td>
<td>0.13</td>
<td>0.86</td>
</tr>
<tr>
<td>Children</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>None dependent (n = 36)</td>
<td>0.89</td>
<td>0.12</td>
<td>0.88</td>
</tr>
<tr>
<td>One or more dependent (n = 12)</td>
<td>0.86</td>
<td>0.14</td>
<td>0.82</td>
</tr>
</tbody>
</table>

### Measures employed in the more detailed arm of the study

#### Medical and demographic characteristics

Current inflammation (C-reactive protein; CRP), duration of RA (since diagnosis), use of disease-modifying anti-rheumatic drugs (DMARDs) and co-morbid medical conditions were noted from hospital records. Age, sex, education, employment, marital status and number of children living at home were recorded during the interview.

#### Knowledge about RA

The Patient Knowledge Questionnaire for RA (PKQ) [21] consists of 16 multiple choice questions on general knowledge of RA (such as aetiology and symptoms), medications, exercise and joint protection.

#### Medication beliefs

The Beliefs about Medicines Questionnaire (BMQ) [31] consists of four validated subscales with items responded to on a five-point Likert scale: general overuse, general harm, specific necessity and specific concerns [31] that have been employed among British people with RA [22, 32].

#### Desirable responding

The self-deceptive enhancement section of the Balanced Inventory of Desirable Responding (BIDR) [23] consists of 20 statements, half of which are accepted to be false (e.g. I always know why I like things) and the other half true (e.g. When my emotions are aroused, it biases my thinking), answered on how true they are believed to be from 1 = ‘not true’ to 7 = ‘very true’. The BIDR is scored as the sum of the true items rated not true (1 or 2) and the false items rated very true (6 or 7) [23].

#### Social support

The Social Support Survey (SSS) [33] uses a four-point scoring system and assesses the perceived level of helpfulness of the support received from various people or groups (spouse, children, other relatives and friends) and has been extended to cover medical support (from the participants’ GP, practice nurses, hospital doctors and hospital nurses) among RA patients [34].

#### Optimism

The Life Orientation Test (LOT) [35] assesses optimism on a unidimensional scale from optimism to pessimism using a five-point Likert scoring system [35].

#### Depression and anxiety

The Hospital Anxiety and Depression Scale (HADS) [36] was employed and consists of 14 items (seven for depression and seven for anxiety) with various four-point scoring systems, and has a period of reference of the past week. Item correlation and comparison with psychiatric ratings showed both subscales to be discriminately valid [36]. The cut-off score for probable depression or anxiety of 10 [36] was also employed.

### Statistical analyses

Patients’ HUs derived from the SG, TTO and EQ-5D were examined for normality before parametric inferential statistical tests were employed. Differences across the HU scores were compared using repeated measures ANOVAs with Bonferroni adjusted post-hoc paired t-tests. Associations between the HU scores derived from the SG, TTO and EQ-5D with functional disability (HAQ) and pain were computed using Pearson’s correlation, \(r\). In the sub sample with more measures, correlations were also computed between the HU scores with the continuous demographic (age and years of education), medical (duration of RA, inflammation, pain and functional disability) and psychosocial variables (knowledge of RA, medication beliefs, desirable responding, social support, optimism, depression and anxiety).
For categorical demographic (sex, having secondary education, being in current paid employment, being married and having dependent children) and medical factors (RA duration group, taking DMARDs, having co-morbidity and clinical HADS scores for depression and anxiety), the HUs were compared between the two resultant subsamples using independent t-tests.

Results

Full sample

Comparison of HU across measures. TTO and EQ-5D scores had acceptable normal distributions (kurtosis = 0.30 and -0.75; skewness = -1.12 and -0.39, respectively), but SG did not (kurtosis = 4.21; skewness = -1.75). One participant had a SG of 0.1 that was a statistical outlier (z = -5.02); this score was removed from all analyses and the normality became acceptable (kurtosis = 0.90; skewness = -1.23; no further statistical outliers with |z| > 4 existed). HUs derived from SG and TTO produced similar mean scores: 0.88 (SD 0.14; range 0.4–1) and 0.86 (SD 0.19; range 0.25–1), respectively. EQ-5D gave a lower HU, with a mean score of 0.52, and a much wider range (SD 0.30; range -0.24 to 1). Within participants, HU differed across the three measures (P < 0.001). Scores derived from EQ-5D was significantly higher than from SG (HU difference, 0.36; adjusted P < 0.001) or TTO (HU difference, 0.34; adjusted P < 0.001).

There were moderate correlations between HU derived through the three different measures: SG scores correlated with TTO scores (r = 0.44, P < 0.001); EQ-5D scores correlated less with SG (r = 0.27, P < 0.001) but moderately with TTO (r = 0.48, P < 0.001).

Relationships of HU with medical factors. In the full sample, SG correlated significantly with HAQ functional disability (r = -0.28, P < 0.001) but not VAS pain (r = -0.08, P = 0.32). TTO and EQ-5D correlated significantly with HAQ functional disability (r = -0.31 and -0.67, respectively, both P < 0.001) and VAS pain (r = -0.19, P < 0.05, and -0.36, P < 0.001). Splitting HAQ functional disability into six score levels graphically demonstrates the stronger linear relation with EQ-5D (Figure 1).

Subsample with additional data

Confirmation of comparison in HU across measures. The subsample of 48 patients in the more detailed study gave very similar results to the full sample of 142 participants. SG, TTO and EQ-5D scores were acceptably normally distributed (kurtosis = 1.02, 1.21 and -0.62; skewness = -1.23, -1.43 and -0.53, respectively). Independent sample t-tests showed no difference in HAQ functional disability (P = 0.18) or VAS pain (P = 0.37) between this subsample and the remaining sample. HUs derived from SG and TTO produced similar mean scores of 0.88 (S.D. 0.13) and 0.87 (S.D. 0.19), respectively (both with a range of 0.5–1). EQ-5D gave a lower mean (0.58), and wider range (0.06–1; S.D. 0.29) than SG or TTO, but was higher in this subsample than in the remaining sample (P < 0.05). Within participants, HU differed similarly across the three measures (P < 0.001) in this subsample. Scores derived from EQ-5D were similarly significantly higher than SG (HU difference, 0.30; adjusted P < 0.001) or TTO (HU difference, 0.29; adjusted P < 0.001).

There were slightly stronger correlations between HUs derived through the three different measures in this subsample: SG scores correlated with TTO scores (r = 0.48, P < 0.001). EQ-5D scores correlated less with SG (r = 0.39, P < 0.01) than TTO (r = 0.62, P < 0.001).

Relationships of HU with medical factors. In the subsample HU scores from SG, TTO and EQ-5D all correlated significantly (negatively) with pain and functional disability (Table 1; at least P < 0.01). HU did not correlate with CRP (Table 1), and no differences were seen in HU by the presence or absence of physical co-morbidity or use of DMARDs (Table 3). Patients with early RA (<6 months since diagnosis) had lower HU on the EQ-5D than those with established disease (>2 yr since diagnosis) (Table 3; P < 0.05).

Relationships of HU with demographic factors. There was only one association between demographic characteristics and HUs: participants in paid employment had significantly higher HU on the EQ-5D than participants who were out of employment (Table 2; P < 0.001). No significant associations were seen between HU and age, education, employment, marital status or presence of children at home (Tables 1 and 2).

Relationships of HU with psychosocial factors. Depression correlated significantly with EQ-5D and TTO to a lesser extent; anxiety only correlated significantly with EQ-5D (Table 1). HU on the EQ-5D was significantly lower among those with clinical depression compared with those without (Table 3; P < 0.01) and among those with clinical anxiety compared with those without mental health problems (Table 3; P < 0.001).
Discussion

Accurate comparison of the cost-effectiveness of different treatment interventions for RA assumes consistent computation of HU. However, large discrepancies have been reported in the literature in this and other patient groups, particularly those with chronic illness. Despite this, few studies have compared HU scores elicited by different methods among people with RA. In the present study, we compared HU derived by two direct instruments (SG, TTO) and one indirect measure (EQ-5D) across 142 RA patients. HU derived from the SG and TTO was significantly higher than that derived from EQ-5D. All HU scores correlated with health status (pain and functional disability), although direct instruments only demonstrated moderate correlations, consistent with previous research [37–39].

Differences between HU across different indirect measures may reflect differences in the underlying constructs [11, 13]. The moderate correlation between direct HU and health status could be due to HU correlating with other aspects of the patient’s life, such as demographic and psychosocial factors. This is the first study to examine this suggestion in detail among people with RA. However, we found little evidence that these factors caused an alteration in risk attitude: minimal relationships with directly elicited HU were apparent in the present study.

None of the demographic factors related to HU derived by direct instruments. There was a single relationship with the EQ-5D: participants in paid employment had significantly higher HU than unemployed participants. This is consistent with findings in previous research where RA patients in employment report lower depression [40] and functional disability [41], which form the majority of the five indices of the EQ-5D. While the present sample contained a relatively small proportion of participants in employment, a considerable difference was observed from people in and out of employment, suggesting that employment status may have an influence on HU and that in turn could have an effect on the health economics of RA.

Patients with established RA also reported higher HU on the EQ-5D consistent with studies showing that such patients have higher acceptance of their illness [42] and lower depression [43]. The relationship of duration of RA and employment status with EQ-5D scores may partly explain the lower HUs elicited by this indirect method, and should be considered in future applications of this instrument among people with RA.

Indirect measures of HU like the EQ-5D explicitly incorporate psychosocial factors. This study is the first to examine whether during direct elicitation of HU, patients also consider psychosocial factors in their health state construct. While SG related only to functional disability and pain, TTO and EQ-5D also correlated with depression, and EQ-5D further correlated with anxiety. The relationship of functional disability, pain, depression and anxiety with EQ-5D scores is anticipated, as similar items are included in the questionnaire. The effect of depression on HU elicited by TTO, but not SG, suggests that these direct instruments represent subtly different health state constructs. This may explain why SG elicits higher HU than TTO and both are higher than EQ-5D in RA.

An alternative explanation of the higher HUs elicited by direct methods (SG and TTO) is that these instruments are inappropriate for most people with chronic illnesses like RA, and only suitable for patients awaiting surgery or with terminal conditions [44]. In RA, the risk of dying after taking the hypothetical medication may appear too unrealistic a gamble to consider. Although there are life-threatening risks of RA treatments, these are very low [3]. The disease itself is also not perceived as life-threatening, despite evidence of its association with potentially fatal co-morbidities and thus reduced life expectancy [45]. Hence, these two direct means of assessing HU are limited for economic evaluations of treatments for most RA patients, and this has implications for clinical trials including CEA. This is consistent with other studies using direct instruments to assess HU in conditions that are not life-threatening. For instance, in patients with intermittent claudication, mean HU derived from SG and TTO has been found to be 0.91 and 0.80, respectively [46]. High baseline HU scores will make any influence of treatment interventions on HU difficult to assess.

Direct HU instruments may become useful in RA if they are modified so that patients compare more moderate health states [47], for instance different levels of functional disability (as highlighted in Figure 1) rather than the extremes of perfect health or death. This has been described as the method of chaining [4–6] and has been implemented for determining HU for patients with oesophageal cancer [48] and in patients with Crohn’s disease [49], another chronic but generally non-terminal illness. Using chaining, the HU of an alteration in functional disability could be determined. However, there are theoretical and administrative problems with chaining because the SG appears to be internally inconsistent across levels of outcome offered [50] and may be even harder to comprehend than the original procedure, which has been suggested to be too complex [14].

The present study found moderate correlations between health status (functional disability and pain) and HU scores derived by the direct instruments SG and TTO, but stronger correlation with the score derived from the indirect measure, EQ-5D. These differences present a problem for comparing different CEA of treatments of RA if HU has been derived by different methods [37–39]. Previous research has suggested the psychosocial or demographic characteristics of patients may lead to the moderate correlation between HU and health status. However, this claim has not been previously examined. This study provides evidence that these moderate associations between HU scores from the SG and TTO and health status are not explained by demographic or psychosocial factors. Instead we propose that the choices offered on the SG and TTO are unrealistic for patients with chronic illness who do not perceive their condition or its treatment to be life-threatening. It is likely that this finding is generalizable to other rheumatic conditions, but it should be addressed in larger and wider future studies.

This study provides evidence that, in their present form, direct measures of HU may be inappropriate for the weighting of CEA of treatments for RA. Population-based indirect measures like the EQ-5D are less complex and more reflective of other standard measures of health status among RA patients.

Acknowledgements

We would like to thank the patients who gave of their time to participate in this study. This research was funded by the Research & Development Directorate and the Department of Rheumatology of the Dudley Group of Hospitals NHS Trust, UK, the Interdisciplinary Research Fund and the School of Psychology of the University of Birmingham, UK, and an unrestricted educational grant from Amgen Ltd. The Department of Rheumatology of the Dudley Group of Hospitals NHS Trust was in receipt of an Arthritis Research Campaign Integrated Clinical Arthritis Centre grant.

The authors have declared no conflicts of interest.
References


