Article:

Yang, Y., Brazier, J. and Tsuchiya, A. (2014) Effect of adding a sleep dimension to the EQ-5D descriptive system: A "bolt-on" experiment. Medical Decision Making , 34 (1). pp. 42-53. ISSN 0272-989X

https://doi.org/10.1177/0272989X13480428

The effect of adding a 'sleep' dimension to the EQ-5D descriptive system – a 'bolt-on' experiment

(Running head: EQ-5D+Sleep)

Yaling Yang (PhD)[1], John Brazier (PhD)[2], Aki Tsuchiya (PhD) [2 3]

[1]Health Economics Research Group, Brunel University, UK

[2] Health Economics and Decision Making, School of Health and Related Research, The University of Sheffield, UK

[3] Department of Economics, The University of Sheffield, UK

Key words: EQ-5D, add-on, sleep, health state valuation, QALYs

Corresponding author: Dr. Yaling Yang

Address: Health Economics Research Group (HERG)

Brunel University

Uxbridge, Middlesex, UB8 3PH

E-mail: yaling.yang@brunel.ac.uk

(4937 words. Four tables.)

ABSTRACT

Background The generic preference-based measures (GPBMs) have been widely used to obtain health utility scores for calculating Quality Adjusted Life Years (QALYs) for economic evaluations. It has been recognized that GPBMs may miss relevant or important dimensions for some specific conditions. **Objective** The objective of this study is to explore the effect of extending the current EQ-5D descriptive system by adding a sleep dimension. **Methods** A new instrument, "EQ-5D+Sleep", is proposed by adding a sleep dimension to the EQ-5D. Based on an orthogonal design, 18 EQ-5D+Sleep states and EQ-5D states were selected and a valuation study was undertaken interviewing 160 members of the generic public in South Yorkshire using time trade-off (TTO). Econometric models have been fitted to the data. Two null hypotheses were tested: 1) the coefficient for the sleep dimension is not significant; and 2) the inclusion of the sleep dimension has no impact on the way people value the other dimensions so that the coefficients of the original dimensions levels remain unaffected. **Results and conclusions** The results support these two null hypotheses. There seems to be no benefit to adding a sleep dimension to the EQ-5D. Research is required to explore the methodology of adding on dimensions to existing descriptive systems of health.

Key words: EQ-5D, add-on, sleep, health state valuation, QALYs

(207 words)

1. INTRODUCTION

Quality Adjusted Life Years (QALYs) have become a widely used measure of health in economic evaluations to inform decisions regarding different health technologies and interventions. A common approach is to use one of the generic preference-based measures (GPBMs) of health to obtain health utility scores for the calculation of QALYs. It has been claimed that GPBMs are applicable to interventions across different patient groups, and consequently ensure comparability between interventions across different medical conditions and treatments. [1-3]

Yet, GPBMs may not be applicable to all interventions across all patient groups. A key concern is that GPBMs may miss relevant or important dimensions for some specific conditions. Generic measures, including those that are preference-based, are intended to cover all important dimensions of health[1]. In practice, however, it is not possible for a measure to address all relevant dimensions of health in their descriptive system to ensure full coverage. For preference-based measures requiring valuation, this is a more obvious problem due to the limited amount of information (e.g. number of dimensions of a measure) individuals are able to handle for valuation. The widely used GPBMs, such as EQ-5D,[4] SF-6D [5] and Health Utility Index 3 (HUI3) [6] typically contain 5 to 8 dimensions to cover the core (but not all) aspects of health. It has been shown that the EQ-5D descriptive system, for example, is insensitive to changes in conditions such as hearing impairment

[7] and visual problems [8;9] and this may be due to inadequate coverage of its descriptive system. An inadequate measure may result in a misallocation of resources.

In order to address the limited relevance of GPBMs, an alternative approach would be to develop condition-specific preference-based measures of health (CSPMBs) for calculating QALYs. CSPBMs have been developed for a wide range of conditions, such as asthma,[10] sexual diseases,[11] and bladder diseases. [12;13] These condition specific measures have raised fundamental concerns as to whether they can be used to make comparisons between interventions for different conditions.[14] CSPBMs have also been criticized because they may fail to pick up side-effects of treatment, and effects on co-morbidities alongside the condition. Furthermore, valuation of CSPBMs may be more prone than GPBMs to 'focusing effects' whereby respondents focus on and exaggerate the importance of the particular problems described and give larger weights to them than they would in the wider context of their overall health. [14]

Given these issues, it seems worthwhile to explore the use of GPBMs as a starting point, and to add items to existing generic measures to improve the relevance to a specific condition. Little attempt has been made to explore such an approach to date. To the best of our knowledge, 3 studies offer comparisons between the EQ-5D against a 6- or 7-dimension instrument that nests it, of which only one is published. First, the original EuroQol instrument can be interpreted as an EQ-5D with an energy/tiredness dimension added , [15] Dowie (1999, Health Economist Study Group conference paper, unpublished manuscript) proposed to add 2 extra dimensions to the EQ-5D questionnaire where one dimension concerned bother from symptoms, and the other dimension concerned bother from treatments. Third, in the same year, Krabbe and his colleagues[16] published an article examining the impact of adding a dimension for cognition to the EQ-5D questionnaire. The resulting EQ-5D+C value set was validated in a population with cognitive impairments by Wolfs and her colleagues. [17]

The main aim of our study was to examine the impact of adding a 'sleep' dimension to the descriptive system of the EQ-5D.

2. WHY ADD 'SLEEP' TO THE EQ-5D?

It has long been observed that the quantity and quality of sleep influence people's self-perceived health, and productivity and performance in society. Sleep problems may include: difficulties falling asleep, waking during the night, waking early and being unable to go back to sleep again and waking in the morning feeling unrefreshed.[18] Clinicians have observed that sleep problems are commonly associated with medical conditions like asthma, or urinary impairment and their treatment. However, sleep problems are also common within the general population. For instance, Groeger et al [18] showed that amongst a representative sample of 2000 British adults, 58% reported sleep problems on one or more nights in the previous week. The corresponding figures for 1010 Americans by gender are 63% for women and 54% for men.[19]

Sleep can have an impact on Subjective Well-being (SWB) where SWB is defined as people's self-reported satisfaction with their own life. For instance, studies have suggested that poorer sleep associated with leads to poorer day-to-day emotional well being [20] and lower global levels of life satisfaction.[21; 22]. These results are in line with the assumption that the effect of sleep matters on people's well-being, and support adding sleep as a dimension to a health related quality of life instrument. In fact, GPBMs such as 15D [23] and the Assessment of Quality of Life (AQoL) instruments,[24] and CSPBMs such as OAB-5D for overactive bladder [12] and AQL-5D for asthma [25] contain a 'sleep' dimension in their descriptive systems. The World Health Organization Quality of Life instrument (WHOQoL-100) also contains 2 sleep-related items. [26]

The EQ-5D was selected as the GPBM to build on because it is the instrument of choice by the National Institute of Health and Clinical Excellence (NICE) in the UK and has been used most widely in economic evaluations. Secondly, it only has 5 dimensions making it easier to add to, compared to the larger descriptive systems like the SF-6D and HUI3 with 6 and 8 dimensions respectively. Furthermore, sleep is not covered by the EQ-5D descriptive system although it was developed as a multi-attribute classification system covering the full health spectrum, indicating physical, mental and social function, to ensure its use across a wide range of health interventions. [4] Given the impact of sleep on people's quality of life, it makes a good candidate to add to the EQ-5D in order to explore the effect of extending the current EQ-5D.

The question is whether sleep has a significant impact on the value of the health states and if so, what the nature of the impact is. Adding a sleep dimension can impact on health state values through significant coefficients for having sleep problems. It may also have an impact on the weight given to the other dimensions of the EQ-5D so that the coefficients for having problems in the original 5 dimensions may be affected. Our research tests these potential impacts using a rigorously designed valuation survey of the general public using an established choice-based valuation technique.

3. METHODS

A valuation study of the EQ-5D+Sleep instrument was conducted using a modified Measurement and Valuation of Health (MVH) Time Trade-off (TTO) protocol. [27] Half the respondents were asked to value hypothetical health states described using the standard EQ-5D instrument without the sleep dimension, the other half to value hypothetical EQ-5D+Sleep states, and their values compared. Econometric models were estimated to predict values for every health state defined by the 2 descriptive systems. The model coefficients for the EQ-5D and the EQ-5D+Sleep were examined and compared to test 2 null hypotheses:

--- the sleep dimension does not have significant impact on health state values (so the coefficients for the sleep dimension are not significant in the EQ-5D+Sleep model); and

--- the inclusion of the sleep dimension in EQ-5D+Sleep has no impact on the way people value the original dimensions of the EQ-5D descriptive system (so that the absolute values of the coefficients of the 5 original dimensions of the EQ-5D+Sleep model are no smaller than the corresponding coefficients of the EQ-5D model).

3.1 INSTRUMENTS AND THE SELECTION OF STATES

***EQ-5D*:** The EQ-5D descriptive system contains 5 dimensions: mobility, self-care, usual activities, pain/discomfort and anxiety/depression. In its most widely used form, each dimension has 3 levels denoting no problems (level 1), moderate problems (level 2) and extreme or severe problems (level 3). This 5-dimension and 3-level health classification system defines a total of 243 (calculated as 3^5) health states. [28] The most widely used scoring algorithm for the EQ-5D was estimated by the MVH group at the University of York using a TTO protocol. [29]

***EQ-5D+Sleep*:** The EQ-5D+Sleep is an extension of the standard EQ-5D descriptive system. A sleep dimension was added to the EQ-5D as the 6th dimension which consists of the following 3 levels to ensure consistency with the existing dimensions:

--- (level 1) I have no problems with sleep

--- (level 2) I have some problems with sleep

--- (level 3) I have extreme problems with sleep

This 6-dimension and 3-level health classification system defines a total of 729 (calculated as 3^6) health states. Each state can be described using a 6-digit code indicating the level at each of the 6 dimensions.

Orthogonal designs generated using SPSS version 15 (SPSS, Inc, an IBM Company, Chicago IL) indicated that for a 5-dimensional 3-level instrument (EQ-5D) and a 6-dimensional 3-level instrument (EQ-5D+Sleep), an additive model can be estimated based on valuations of 18 states each. The best state defined by the instrument (state 11111 for EQ-5D and state 111111 for EQ-5D+Sleep) were included in both sets. As these best states were going to be used as the upper anchors in the TTO valuation task, 17 intermediate states were selected for valuation. For each instrument, the intermediate states were stratified into severity groups based on their total level scores across the dimensions and then randomly allocated to blocks of either 8 or 9 states, resulting in 2 EQ-5D blocks and 2 EQ-5D+Sleep blocks. Then the ‘pits’ or the worst possible state (33333 for EQ-5D and 333333 for EQ-5D-Sleep) of each instrument was added to each block.

Among the health states suggested by the orthogonal design, there were 6 ‘matched health state pairs’ across the EQ-5D and EQ-5D+Sleep states, where a matched pair contained an EQ-5D state and a corresponding EQ-

5D+Sleep state consisting of the same EQ-5D profile plus any given level of the extra sleep dimension. The 6 matched health state pairs were: 11233 vs. 112331, 22232 vs. 222321, 12312 vs. 123122, 33132 vs. 331322, 31211 vs. 312113, and 33333 vs. 333333. It can be seen that 2 states add sleep at level 1, 2 states add sleep at level 2, and a further 2 states add sleep at level 3. The health state values of these matched state pairs can be compared to determine the impact of adding the sleep dimension.

3.2 THE INTERVIEWS

The main part of the study consisted of a valuation survey of selected EQ-5D states and EQ-5D+Sleep states among members of the public in South Yorkshire, UK (adults over 18 years old). The first stage of recruitment was to select a sample of streets within South Yorkshire using the ADF Names and Numbers database (ADF Software Limited, Ramsey, UK, available from http://www.afd.co.uk/product_namesandnumbers.asp) which provides access to names and addresses for over 39 million people in the UK. The sampling aimed to achieve a good spread across age, gender, ethnicity and social class. An information sheet was then sent to all household addresses on the sampled streets explaining the project in plain language and inviting their participation. In the next stage, an interviewer knocked on the doors of randomly selected households to obtain the resident's consent to participate, and either interviewed the resident immediately or arranged a convenient time to revisit for the interview. There was no limit to the number of interviews conducted in a selected household.

There is little guidance available on how many observations are required per health state in order to model health state valuations. Given the budget constraints, the sample size for the valuation survey was limited to 160 interviews. The respondents were randomly divided into 4 groups of 40: and each group was assigned one of the 4 health state blocks explained above. In this way, each intermediate state was valued 40 times, which is reasonable compared to the average of 15 times per state for the SF-6D valuation study [(5)] or the average of 24 times for the UK HUI survey [(30)] given that the classification systems of SF-6D and HUI are much larger than that of EQ-5D or EQ-5D+Sleep. The pits states were valued 80 times each.

The interviews were based largely on the MVH valuation protocol, [(29)] and included self-reported health status using either EQ-5D or EQ-5D+Sleep (the instrument they are valuing), ranking of hypothetical states, and a valuation exercise using Time Trade Off [(27)], followed by personal background questions. Those respondents valuing EQ-5D states were given an extra question using the sleep dimension of the EQ-5D+Sleep as part of the background questions at the end of the interview.

3.3 THE ANALYSIS

The background characteristics of the 2 groups of respondents who valued the EQ-5D and the EQ-5D+Sleep descriptive systems were compared. A Chi-*s*quare test was performed to examine whether there was any difference between the 2 samples in terms of age group, gender, or self-reported health status. For ordered variables (e.g. age groups ’18-25’, ’26-35’, ‘36-45’ etc, and self-reported EQ-5D dimensions from ‘1’ of ‘no problems’, ‘2’ of moderate problems and ‘3’ of ‘extreme problems’), the Chi-square Gamma statistic was undertaken to make adequate use of the relevant information.

The TTO valuations derived for the health states defined by the 2 descriptive systems were transformed following Dolan (31) which ensures all health state values are bound between (-1) and (+1). The number of observations, mean transformed TTO values and standard deviations, and maximum and minimum values are reported for the 2 instruments. The *t*-tests were performed to test for significant differences between the 6 matched health state pairs across the 2 instruments.

The next stage was to model the health state values of EQ-5D and EQ-5D+Sleep on the basis of the valuation data directly obtained from the survey. The main purpose of modelling here is to compare the model coefficients to test the 2 null hypotheses above. STATA version 9 (StataCorp, College Station, TX) was used for all regression analysis.

The general model is defined as:

$$y_{ij} = g(\beta'\mathbf{x}_{\delta\lambda} + \theta'\mathbf{r}_{\delta\lambda} + \delta'\mathbf{z}_j) + \varepsilon_{ij} \quad \textbf{(1)}$$

Where:

y_{ij}: transformedTTO scores for health state *i* valued by respondent *j*

i = 1, 2, …, n represents individual health states

j = 1, 2, …, m represents respondents

g: a function specifying the appropriate functional form

x: a vector of binary dummy variables for each level λ of dimension δ of the classification of EQ-5D or EQ-5D+Sleep

r: a vector of terms to account for interactions between the levels of different dimensions of EQ-5D or EQ-5D+Sleep

z: a vector of personal characteristics including respondent's gender, age, self-reported health (individual EQ-5D dimensions and sleep), education and household status.

ε_{ij}: an error term whose autocorrelation structure and distributional properties depend on the assumptions underlying the particular regression model used.

A range of models were considered and the ultimate choice of model specification was based upon the features of the valuation data obtained. An ordinary least squares (OLS) regression using individual level TTO values would assume that each health state value is independent, ignoring the fact that multiple health state valuations have been given by the same respondent so that health state values may be clustered by respondents. A potentially better specification is the one-way error components random effects (RE) model which takes account of the clustering of data by respondents and allows for the fact that the error term may not be independent of the respondent – it separates out error terms both within and between respondents. The RE model also assumes that the error term for any health state valuation by an individual is random. The RE model specification has previously been successfully estimated and chosen as 'the best model' for EQ-5D valuation data obtained from the UK general population [29] and in several other cases. [32;33] Therefore for the current study, a one-way error components RE model using individual level data was estimated for each of the 2 descriptive systems using maximum likelihood estimation.

In order to ensure that the coefficients of the level 2 and level 3 dummies would have a more intuitive interpretation, decrements in utility from full health are modelled, by using the difference between 1 and the TTO value as the dependent variable. If a respondent values a health state as worse than being dead, the associated disutility would be larger than 1. Thus, the level 2 and level 3 coefficients are expected to be positive, and with level 3 having a larger coefficient.

Interaction terms between all the different levels of all the dimensions of EQ-5D or EQ-5D+Sleep were not feasible due to the relatively small sample size. Further, the health state selection was based on a linear additive model. Therefore, only Interaction terms between the sleep and the EQ-5D dimensions were examined when the EQ-5D dimensions showed significant change of coefficients after the inclusion of the sleep dimension (see next paragraph).

After the model estimation, the coefficients of the EQ-5D and EQ-5D+Sleep models were compared to test the 2 null hypotheses presented above. The first hypothesis was straightforward, while the second hypothesis was examined by means of a comparison between the modelled coefficients for the 5 original EQ-5D dimensions and those for EQ-5D+Sleep using a series of z-tests for the 10 corresponding beta coefficients (see equation 2) with a 0.10 significance level [34;35].

$$z_{\delta\lambda} = \frac{\beta^{E}_{\delta\lambda} - \beta^{S}_{\delta\lambda}}{\sqrt{(SE^{E}_{\delta\lambda})^2 + (SE^{S}_{\delta\gamma})^2}} \qquad (2)$$

Beta: regression coefficient
SE: standard error of coefficient
E: EQ-5D
S: EQ-5D+sleep

4. RESULTS

4.1 SAMPLE BACKGROUND CHARACTERISTICS

The EQ-5D and EQ-5D+Sleep valuation survey was undertaken between June and September 2007. Nine hundred invitation letters were mailed out. From these, a total of 160 members of the public were successfully interviewed (80 for EQ-5D and 80 for EQ-5D+Sleep) and their data included in subsequent analysis. Recruitment information is reported in Table 1, including a response rate of 51% amongst those eligible (i.e. those at home and suitable for interview).

{Insert Table 1 here}

The personal characteristics of the 2 samples that valued EQ-5D and the

EQ-5D+Sleep are reported in Table 2. In general, the 2 samples were comparable in terms of age, gender, education and social-economic status. The results of the Chi-square test suggested that there was no significant difference between the 2 samples in terms of age, gender and health status (P>0.1). The 2 samples had similar health status as described by the EQ-5D, but more respondents in the EQ-5D sample suffered from moderate sleep problems than in the EQ-5D+Sleep sample (33/80 vs. 20/80;p<0.1).

{Insert Table 2 here}

4.2 TTO HEALTH STATE VALUES

A total of 1512 TTO values were elicited from the 161 respondents in the interview survey with 770 values for the 18 EQ-5D states and 742 values for the 18 EQ-5D+Sleep states. Descriptive statistics are presented in Table 3, where health states are ranked according to their mean TTO values.

{Insert Table 3 here}

Transformed TTO values at the individual level ranged from -0.98 to 1.00. In terms of mean TTO values, EQ-5D values had a range of -0.227 (state 33333) to 0.609 (state 12312), while EQ-5D+Sleep values ranged from -0.233 (state 333333) to 0.764 (state 211223). The standard deviations of EQ-5D states were from 0.35 to 0.63 with an average of 0.52, while for the EQ-5D+Sleep states these ranged from 0.30 to 0.50 with an average of 0.43. Across the 2 instruments, less severe health states with higher values tended to have smaller standard deviations.

A significant difference between the matched EQ-5D and EQ-5D+Sleep states was only found for health states pair 11233 (mean value 0.18) vs. 112331 (mean value 0.49) where 'no problem' in sleep was added to a generally moderate EQ-5D state (p<0.1). However, no statistically significant differences were found for mean values of the other pairs, namely, 22232 (mean value 0.44) vs. 222321(mean value 0.31); 12312 (0.609) vs. 123122 (0.591); 33132 (0.093) vs. 331322 (0.171); 31211 (0.375) vs. 312113 (0.393); and 33333 (-0.227) vs. 333333 (-0.233); where various levels of the sleep dimension were added to relatively moderate or severe EQ-5D states. For those pairs without significant differences, no obvious pattern was found. The implication here seems to be that, in most cases, adding an extra sleep dimension to the EQ-5D system does not change people's values significantly. This was examined further by modelling the data as reported below.

4.3 HEALTH STATES MODELLING

The results of modelling are presented in Table 4, where models (E1) and (S1) are main effects models for EQ-5D and EQ-5D+Sleep respectively, and models (E2) and (S2) include respondents` socio-demographic variables.

{Insert Table 4 here}

The EQ-5D main effects model (E1) had coefficients in agreement with the ordinality of the EQ-5D health state classification. All coefficients had the expected positive sign except for level 2 of self-care but this coefficient was not statistically significant (p>0.1) so it is a weak inconsistency. Overall, 7 out of a total of 10 main effects coefficients in model E1 were statistically significant (p<0.1) with the expected sign. All the coefficients for level 3 (extreme) problems were significant, but for level 2 only usual activity and pain/discomfort coefficients were significant.

In most cases, the EQ-5D+Sleep main effects model (S1) had coefficients in agreement with the ordinality of the EQ-5D+Sleep health state classification. All coefficients had the expected positive sign except for level 2 for anxiety/depression and level 2 for sleep. For level 2 of anxiety/depression, it was not statistically significant (p>0.1) so it is a weak inconsistency. Of the sleep coefficients, only level 2 was significant (p=0.08) coefficient, but with an unexpected sign. Overall, 8 out of the 12 main effects coefficients in model S1 were statistically significant, with the expected sign. The 3 non-significant coefficients are level 2 pain/discomfort, level 2 anxiety/depression, and level 3 sleep.

Comparing the EQ-5D main effects model E1 with the corresponding EQ-5D+Sleep model S1, 7 out of 10 coefficients of the EQ-5D model were significant (p<0.1) with the expected sign, while 8 out of 12 coefficients were significant with the expected sign in the EQ-5D+Sleep model. Within each descriptive system, the regression coefficients were logically ordered, with the exception of the sleep coefficients. Both models showed that extreme problems in the mobility dimension contributed to the most disutility. In terms of the least disutility caused by a level 3 problem, this was the usual activity dimension in the EQ-5D model and the sleep dimension in the EQ-5D+Sleep model. The EQ-5D model had a constant term of 0.281 while the EQ-5D+Sleep model had a smaller constant term of 0.146.

Generally, the coefficients of the 2 models were comparable as their differences were small (e.g. 0.315 vs. 0.307 in the mobility dimension). The exception was the self-care dimension which had a difference larger than 0.1 (0.122 vs. 0.259). The results of z-tests between the corresponding coefficients of the 2 models quantitatively supported this conclusion. Only the coefficients of the self-care dimension were statistically significantly different across the models. For both level 2 and level 3 self-care, the coefficient in the S1 model is larger than the coefficient in the E1 model, indicating that the inclusion of the sleep dimension increased the impact of the self-care dimension on TTO values. Interactions between the self-care and sleep dimensions were examined. Four interaction terms were included in the model (selfcare2 * sleep2, selfcare2 * sleep3, selfcare3 * sleep2, and selfcare3* sleep3), but none of the associated coefficients was statistically significant($P<0.1$).

Respondents' age, gender, education, household status and self-reported health (in EQ-5D dimensions and sleep) were introduced into the EQ-5D and the EQ-5D+Sleep models (models E2 and S2) in order to control for any differences between the 2 samples and capture effects of covariates. The results show that none of these coefficients were significant in the EQ-5D model, and for the EQ-5D+Sleep model, none but 3 (of the 5) age group coefficients were statistically significant. For either model, the inclusion of covariates rarely changed the coefficients of the original models, apart from the constant term of the EQ-5D model which decreased from 0.281($p<0.1$) to -0.423 ($p>0.1$). The constants of both the EQ-5D and the EQ-5D+Sleep models went from statistically significant to non-significant by including covariates.

5 DISCUSSION AND CONCLUSION

Recognising the importance of sleep for people's quality of life, this study attempted to examine the effect of adding a sleep dimension onto the original EQ-5D descriptive system. The evidence did not support the hypothesis that the sleep dimension makes a significant impact on the values people place on the EQ-5D. First, among the 6 matched health state pairs, adding a sleep dimension did not change the EQ-5D values significantly in most cases. Secondly, when health state values were modelled, sleep was the only dimension with a non-significant level 3 coefficient. Thirdly, adding a sleep dimension did not induce significant changes to the coefficients of most EQ-5D dimensions. The conclusion holds true when respondents` socio-demographic characteristics were introduced into the models.

These findings were not expected, as 'sleep' was found to be important in the well-being literature. However, the results above need not be regarded as incompatible with the findings in the well-being literature. First, people may not realise the impact of sleep problems on well-being, and thus may not regard having sleep problems as something worth giving up survival for. Alternatively, even if people realise the impact of sleep problems on well-being, if they think the impact is already captured adequately by the EQ-5D, then there may be no further residual impact worth trading survival for. For example, impaired sleep may have been reflected in impaired mobility (e.g., poor balance/reflexes), impaired usual activity (e.g. slow response), and anxiety/depression problems. Qualitative research could be a useful way to explore these issues. On the other hand, the current findings echo the EuroQol Group's early findings using version of their instrument consisting 6 dimensions: the non-significant 'energy' dimension was eventually dropped to form the current EQ-5D instrument. [15] Sleep may strongly correlate with energy. At the same time, the SF-6D [5] contains a 'vitality' dimension described as 'having a lot of energy all of the time / most of the time/ etc' and while the coefficients of this vitality dimension in the SF-6D model are significant, they are smaller in size compared to the other dimensions.

There were 2 exceptions regarding the general non-significant impact of the sleep dimension. One was the matched health state pair 11233 vs. 11233<u>1</u> where specifying 'no sleep problem' resulted in the health state value significantly increasing from 0.179 (11233) to 0.486 (11233<u>1</u>). It may be that people valued the ability to sleep well under extreme pain/discomfort and anxiety/depression, because problem-free sleep would be a form of relief from pain/discomfort and anxiety/depression. Alternatively, respondents may have imagined poor sleep would result from the original EQ-5D state and so adding level 1 for sleep improved the state. However, when level 1 sleep was added to EQ-5D state 22232 (e.g. reflecting extreme pain/discomfort and moderate problems in other dimensions) to form EQ-5D+Sleep state 222321, the 2 states were not found to be statistically significantly different from each other.

Another exception was found via model estimation. While most coefficients are unaffected, the self-care dimension exhibited significantly different coefficients with or without the sleep dimension: and this pattern remained when respondents' socio-demographic variables were introduced to the models. It is possible that interactions between the self-care and sleep dimensions exist. This was tested by creating interaction terms between these 2 attributes (self-care × sleep), but none of the coefficients was statistically significant. However, it should be noted that the study design only allows the estimation of additive models and possibly a larger valuation study with more states is needed for further examination of interactions.

This finding also provides evidence to challenge the assumption that the impact of different dimensions on preferences is additive. If the assumption holds, inclusion or exclusion of a sleep dimension (or any dimension) should lead to no significant change in the coefficients of the other dimensions in the classification. The non-additivity between dimensions was also observed in a recent study of adding on a pain dimension to an asthma-specific utility measure, the AQL-5D.[36] The possibility of non-additivity of dimensions creates a further challenge in the development of the add-on approach because if the relationship between dimensions of health is not additive, then it would be problematic to have a core value set of a preference-based measure (such as EQ-5D) and then to simply add relevant dimensions and their associated decrements on to it. It may be possible to use multiplicative interaction terms to solve this problem but this requires a larger design and would have to be repeated for each potential dimension to be added .

While a simple comparison of values across health states with and without the extra dimension gives an indication of whether this has an impact, in order to conduct a systematic comparison across the 2 descriptive systems and their values, there needs to be a statistically designed survey with modelling to estimate the impact of the additional dimension through its coefficients and those of the other dimensions. The main contribution of this study has been in showing how add-on studies might be performed. Previously in the literature, no valuation study was undertaken in Dowie's work, and the impact of the 2 dimensions on EQ-5D values was ultimately determined through an arbitrary estimation (5% or 10% reduction of the EQ-5D tariff for moderate or extreme bother respectively, for each additional dimension). A problem shared by the testing of the energy dimension in the original EuroQol Instrument and the testing of the cognition dimension in the EQ-5D+C instrument is that both studies used VAS to value the selected health states, rather than a choice-based technique (i.e. Time Trade Off or Standard Gamble); and the health states selected for valuation were not based on a statistical design so there was no formal estimation of the impact on the overall utility function. Furthermore the only published valuation study for an add-on, for the EQ-5D+C, was undertaken using a convenience sample of academic and managerial staff at a university and so is not likely to be representative of the general population. Therefore, our study is the only one to date that has examined the impact of adding a dimension to a generic measure, based on values obtained from members of the general population using a choice-based technique and one that examined the implications for the overall utility function.

This study has a number of limitations. The first is the choice of the sleep dimension. Specifically, problems with sleep could be related to sleep disturbance or sleep-related impairment and these could be considered distinct. Qualitative interviews with members of the general public would help to elucidate this but was not undertaken in the current study. Second, the sample size of the current study was relatively small for estimating a full model for valuing health states. This may limit the study's ability to detect important differences. Nevertheless, these limitations cannot undermine the methodological contribution of the study to the development of add-ons.

For future add-on studies, it is important to examine and choose the dimension(s) to be added to a measure with care. A literature review on the performance of GPBMs in different conditions may suggest possible missing dimensions to add. On the basis of systematic reviews of validity and responsiveness of EQ-5D, vision was suggested as a potential candidate for future add-on studies [(8)] Mapping functions between (non-preference based) condition-specific measures onto GPBMs may also provide information on candidate dimensions to add. Those dimensions that perform poorly in mapping regressions are in effect missing from the GPMBs, and therefore can be regarded as potential candidates for add-ons. Finally, qualitative interviews of patients and/or clinicians could be used to identify possible candidates of add-on dimensions.

Because of its brevity, EQ-5D may miss important dimension(s) of health, and it has been argued that the add-on approach may to some extent offer a solution. This study provides the most robust empirical test of the add-on approach to date through TTO valuations with and without the add-on using econometric models to examine the impact of the extra dimension on the overall preference function for EQ-5D. This rigorous approach needs to be extended to other possible candidate dimensions to add Larger sets of health states and respondent samples are necessary to ensure sufficient power to examine interactions between health dimensions, especially between the original and the added dimensions.

Acknowledgements*:* Ethical approval of this study was given by the School of Health and Related Research (ScHARR) Ethics Committee at the University of Sheffield. We would like to thank Jason Leman and his colleagues at Sheffield Hallam University for undertaking the valuation survey. We are grateful to the survey participants for their contributions. We thank for useful comments from discussant Nancy Devlin and participants of Health Economist Study Group (HESG) conference in January 2008. We are also grateful for helpful comments from the 3 anonymous reviewers for Medical

Decision Making. The usual disclaimers apply. John Brazier and Aki Tsuchiya are members of the EuroQol Group, but this study was carried out independently of the Group.

Reference List

(1) Dowie J. Decision validity should determine whether a generic or condition-specific HRQOL measure is used in health care decisions. Health Economics 2002;11(1):1-8.

(2) National Institute of Health and clincal Excellence (NICE). NICE Guide to the Methods of Technology Appraisal. NICE, London . 2004.

(3) National Institute of Health and Clinical Excellence (NICE). NICE Guide to the Methods of Technology Appraisal. NICE, London 2008.

(4) Brooks R. EuroQol: the current state of play. Health policy 1996;37(1):53-72.

(5) Brazier J, Roberts J, Deverill M. The estimation of a preference-based measure of health from the SF-36. Journal of Health Economics 2002;21(2):271-92.

(6) Feeny D, Furlong W, Torrance GW, Goldsmith CH, Zhu Z, DePauw S, et al. Multiattribute and single-attribute utility functions for the health utilities index mark 3 system. MED CARE 2002;40(2):113-28.

(7) Grutters JPC, Joore MA, van der Horst F, Verschuure H, Dreschler WA, Anteunis LJC. Choosing between measures: comparison of EQ-5D, HUI2 and HUI3 in persons with hearing complaints. Quality of Life Research 2007;16(8):1439-49.

(8) Tosh J, Brazier J, Evans P, Longworth L. A review of generic preference-based measures of health-related quality of life in visual disorders. Value in Health 2011.

(9) Espallargues M, Czoski-Murray CJ, Bansback NJ, Carlton J, Lewis GM, Hughes LA, et al. The impact of age-related macular degeneration on health status utility values. Investigative ophthalmology & visual science 2005;46(11):4016.

(10) Revicki DA, Leidy NK, Brennan-Diemer F, Thompson C, Togias A. Development and preliminary validation of the multiattribute Rhinitis Symptom Utility Index. Quality of Life Research 1998;7(8):693-702.

(11) Torrance GW, Keresteci MA, Casey RW, Rosner AJ, Ryan N, Breton MC. Development and initial validation of a new preference-based disease-specific health-related quality of life instrument for erectile function. Quality of Life Research 2004;13(2):349-59.

(12) Yang Y, Brazier J, Tsuchiya A, Coyne K. Estimating a PreferenceΓÇÉBased Single Index from the Overactive Bladder Questionnaire. VALUE HEALTH 2009;12(1):159-66.

(13) Brazier J, Czoski-Murray C, Roberts J, Brown M, Symonds T, Kelleher C. Estimation of a preference-based index from a condition-specific measure: the King's Health Questionnaire. MED DECIS MAKING 2008;28(1):113.

(14) Brazier J, Rowen D, Mavranezoui I, Tsuchiya A, Young T, Yang Y. Developping and testing methods for deriving preference-basd measures of health from condtion specific measures (and other patient based measures of outcome). Health Technology Assessment. Forthcoming. 2012.

(15) Gudex C. Are we lacking a dimension of energy in the Euroqol instrument? Paper presented at the 8th Pleanary Meeting of the EuroQol Group, Lund, Sween. 17-10-1991.

(16) Krabbe PFM, Stouthard MEA, Essink-Bot ML, Bonsel GJ. The effect of adding a cognitive dimension to the EuroQol multiattribute health-status classification system. Journal of Clinical Epidemiology 1999;52(4):293-301.

(17) Wolfs CAG, Dirksen CD, Kessels A, Willems D, Verhey FRJ, Severens JL. Performance of the EQ-5 D and the EQ-5 D+ C in elderly patients with cognitive impairments. Health and Quality of Life Outcomes 2007;5(1):33.

(18) Groeger JA, Zijlstra FRH, Dijk DJ. Sleep quantity, sleep difficulties and their perceived consequences in a representative sample of some 2000 British adults. Journal of Sleep Research 2004;13(4):359-71.

(19) National sleep foundation. Executive summay of the 2003 sleep in American poll. Available from: http://www.sleepfoundation.org/article/sleep-america-polls/2003-sleep-and-aging . 2003.

(20) Kahneman D, Krueger AB, Schkade DA, Schwarz N, Stone AA. A survey method for characterizing daily life experience: The day reconstruction method. Science 2004;306(5702):1776.

(21) Carbonell A. Income and well-being: an empirical analysis of the comparison income effect. Journal of Public Economics 2005;89(5-6):997-1019.

(22) Yokoyama E, Saito Y, Kaneita Y, Ohida T, Harano S, Tamaki T, et al. Association between subjective well-being and sleep among the elderly in Japan. Sleep Medicine 2008;9(2):157-64.

(23) Sintonen H. The 15D-measure of health-related quality of life. II. Feasibility, reliability and validity of its valuation system. National Centre for Health Program Evaluation, Melbourne, Australia 1995.

(24) Hawthorne G, Richardson J, Osborne R. The Assessment of Quality of Life (AQoL) instrument: a psychometric measure of health-related quality of life. Quality of Life Research 1999;8(3):209-24.

(25) Yang Y, Brazier JE, Tsuchiya A, Young TA. Estimating a preference-based index for a 5-dimensional health state classification for asthma derived from the asthma quality of life questionnaire. MED DECIS MAKING 2011;31(2):281.

(26) World Health Organization (WHO), Geneva. World Health Organization Quality of Life Instrument (WHOWoL-100). 1995.

(27) Gudex C. *Time trade-off user manual: Props and self-completion Methods*. Centre for health Economics, University of York, York. 1994.

(28) Kind P. The EuroQol instrument: an index of health-related quality of life. Quality of life and pharmacoeconomics in clinical trials 1996;191-201.

(29) Dolan P. Modeling valuations for EuroQol health states. MED CARE 1997;1095-108.

(30) McCabe CJ, Stevens KJ, Brazier JE. Utility scores for the Health Utilities Index Mark 2: an empirical assessment of alternative mapping functions. MED CARE 2005;43(6):627.

(31) Dolan P. Modeling valuations for EuroQol health states. MED CARE 1997;1095-108.

(32) Tsuchiya A, Ikeda S, Ikegami N, Nishimura S, Sakai I, Fukuda T, et al. Estimating an EQΓÇÉ5D population value set: the case of Japan. Health Economics 2002;11(4):341-53.

(33) Shaw JW, Johnson JA, Coons SJ. US valuation of the EQ-5D health states: development and testing of the D1 valuation model. MED CARE 2005;43(3):203.

(34) Paternoster R, Brame R, Mazerolle P, Piquero A. Using the correct statistical test for the equality of regression coefficients. Criminology 1998;36:859.

(35) Clogg CC, Petkova E, Haritou A. Statistical methods for comparing regression coefficients between models. The American Journal of Sociology 1995;100(5):1261-93.

(36) Brazier J, Rowen D, Tsuchiya A, Yang Y, Young TA. The impact of adding an extra dimension to a preference-based measure. Social Science & Medicine 2011;73(2):245-53.

Table 1 Recruitment information for the valuation survey

Letters of invitation sent	900
Letters returned from vacant properties	12
Refusals at this stage	0
Properties called where residents were not at home	433
Refusals when interviewer called	154
Respondents unsuitable for interview*	53
Number of other reasons for interview not going ahead†	56
Partially completed interviews	2
Completed interviews	160

*The resident was deemed unsuitable if under 18 years old or they had a disability preventing them from completing the survey.

† Number of other reasons for interview not going ahead' usually means that the interviewer could not carry out an interview when they visited the address, but there was a possibility of carrying out an interview at that address at another time. This could be because the resident stated they could not do the interview on this occasion but could on another visit, or were unsure if they wanted to do the interview and requested the interviewer to call back another time, or the person in the house at the time was not suitable (e.g. a child) but the interviewer might have been able to carry out an interview with another resident of the house at another time.

Response rate1 = successful interviews / basic sample

= 160 / (900 – (addressed vacant + unsuitable respondents))

= 160 / (900– (12 + 53))

= 160 / 835 = 19 %

Response rate2

=successful interviews / total number of eligible respondents

=160 / (successful interviews + total number of respondents refusing to participate)

=160 / (160 + 154) =160 / 314= 51%

Table 2 Personal characteristics of the respondents

		EQ-5D sample (N=81)	EQ-5D+Sleep sample (N=80)
Gender	Female	58	52
	Male	23	28
Age	18-25	7	4
	26-35	13	15
	36-45	15	18
	46-55	18	11
	56-65	16	16
	66-	10	15
Self reported EQ mobility	No problems	64	66
	Moderate problems	17	14
	Extreme problems	0	0
Self reported EQ self care	No problems	80	76
	Moderate problems	1	4
	Extreme problems	0	0
Self reported EQ usual activities	No problems	68	67
	Moderate problems	12	12
	Extreme problems	1	1
Self reported EQ pain/discomfort	No problems	49	51
	Moderate problems	29	23
	Extreme problems	3	5
Self reported EQ anxiety/depression	No problems	66	66
	Moderate problems	14	12
	Extreme problems	1	1
Self-reported sleep*	No problems	41	56
	Moderate problems	33	20
	Extreme problems	3	3
Education after minimum school leaving age	Yes	46	50
	No	35	30
Home ownership	Own home outright or with mortgage	60	58
	Rent from a local authority or private sector	21	22

*: P < 0.1

Table 3 Description of TTO values for EQ-5D and EQ-5D+Sleep states

	Health state	N	Mean	Std. Deviation	Minimum	Maximum
EQ-5D	**33333**	81	-0.23	0.52	-0.98	0.99
	32323	41	0.03	0.61	-0.98	1.00
	33132	40	0.09	0.56	-0.98	1.00
	11233*	41	0.18	0.60	-0.88	1.00
	21331	41	0.24	0.57	-0.88	0.99
	31112	41	0.29	0.63	-0.98	1.00
	32221	40	0.32	0.58	-0.93	1.00
	31211	41	0.38	0.61	-0.98	1.00
	23222	41	0.40	0.51	-0.98	1.00
	13213	40	0.40	0.51	-0.88	1.00
	23311	41	0.43	0.52	-0.88	1.00
	22232	40	0.44	0.51	-0.88	1.00
	22113	41	0.46	0.53	-0.93	1.00
	13121	40	0.54	0.47	-0.93	1.00
	21123	40	0.57	0.42	-0.73	1.00
	12131	40	0.58	0.41	-0.88	1.00
	11322	40	0.60	0.42	-0.98	1.00
	12312	41	0.61	0.35	-0.88	1.00
	Total	770	0.32	0.58	-0.98	1.00
EQ-5D +Sleep	**333333**	78	-0.23	0.45	-0.98	0.63
	331322	40	0.17	0.49	-0.93	0.99
	322232	38	0.23	0.47	-0.98	1.00
	223313	38	0.28	0.48	-0.93	1.00
	222321	39	0.31	0.50	-0.93	1.00
	321111	38	0.39	0.43	-0.93	1.00
	231131	38	0.39	0.48	-0.98	1.00
	312113	38	0.39	0.47	-0.93	1.00
	313221	40	0.40	0.46	-0.83	1.00
	232212	39	0.40	0.50	-0.88	1.00
	121233	38	0.44	0.42	-0.98	1.00
	133211	40	0.46	0.45	-0.73	1.00
	112331*	40	0.49	0.47	-0.83	1.00
	132123	40	0.55	0.45	-0.63	1.00
	123122	38	0.59	0.30	-0.48	1.00
	213132	40	0.60	0.31	-0.53	1.00
	111312	40	0.72	0.30	-0.38	1.00
	211223	40	0.76	0.32	-0.48	1.00
	Total	742	0.38	0.50	-0.98	1.00

Paired health states are in bold.
* Significant difference between the matched health states ($p < 0.1$) using *t*-test

Table 4 Random effects models for EQ-5D and EQ-5D+Sleep health states

Dimensions & levels	Main effects model			Main effects model with covariants	
	EQ-5D (E1)	EQ-5D+Sleep (S1)	Z score (E1) vs.(S1)	EQ-5D (E2)	EQ-5D+Sleep (S2)
Constant	0.281**	0.146**	1.659	-0.423	0.103
Mobility					
2	0.034	0.065**	-0.674	0.026	0.054*
3	0.315**	0.307**	0.185	0.292**	0.297**
Self-care					
2	-0.017	0.141**	-3.038**	-0.010	0.131**
3	0.122**	0.259**	-3.172**	0.133**	0.262**
Usual activity					
2	0.057*	0.073**	-0.364	0.064**	0.072**
3	0.108**	0.137**	-0.658	0.096**	0.137**
Pain/discomfort					
2	0.059*	0.042	0.368	0.075**	0.045
3	0.242**	0.208**	0.761	0.260**	0.209**
Anxiety/depresión					
2	0.008	-0.005	0.292	0.018	0.001
3	0.168**	0.130**	0.910	0.176**	0.127
Sleep					
2	-	-0.049*	-	-	-0.052*
3	-	0.036	-	-	0.036
Female	-	-	-	0.056	-0.105
Age			-		
26-35	-	-	-	-0.019	-0.205
36-45	-	-	-	-0.160	-0.333*
46-55	-	-	-	-0.055	-0.392**
56-65	-	-	-	0.072	-0.379*
66-	-	-	-	0.070	-0.178
Renting home			-	-0.034	-0.078
No education after mínimum School leaving age	-	-	-	-0.170	-0.143
Self-reported morbidity					
2	-	-	-	-0.051	0.024
3	-	-	-	-	-
Self-reported self-care					
2	-	-	-	0.010	-0.089
3	-	-	-	-	-
Self-reported usual activities	-				
2	-	-	-	0.027	0.608
3	-	-	-	0.272	0.469
Self-reported pain/discomfort					
2	-	-	-	0.303	0.265
3	-	-	-	0.194	0.191
Self-report anxiety/depression					
2	-	-	-	0.560	-0.313
3	-	-	-	0.260	-0.302
Self-reported sleep					
2	-	-	-	-0.131	0.096
3	-		-	-0.016	0.118
Adjusted R^2	0.174	0.256	n/a	0.280	0.344

Note : Dependent variable disutility = 1 – tto.
* p<0.1 **p<0.05