



# Mapping between HAQ-DI and EQ-5D-5L in a Chinese patient population

Thomas Patton<sup>1</sup> · Hao Hu<sup>2</sup> · Luan Luan<sup>2</sup> · Keqin Yang<sup>3</sup> · Shu-Chuen Li<sup>2</sup>

Accepted: 21 June 2018 / Published online: 4 July 2018  
© The Author(s) 2018

## Abstract

**Objectives** In order to address the current deficiency of health utility evidence relevant for economic evaluations involving treatments for rheumatoid arthritis (RA) in the Chinese setting, this study aims to develop a mapping algorithm linking the Health Assessment Questionnaire (HAQ) and EQ-5D-5L in a Chinese population of patients with RA.

**Methods** An estimation sample was obtained from a cross-sectional study that collected HAQ, the pain Visual Analogue Scale, and EQ-5D-5L in RA patients in two tertiary referral hospitals in China. Mapping algorithms were derived in this study using two alternative regression methods: the beta regression and a multivariate ordered probit regression. The internal validity of the mapping algorithms was assessed in each case by calculating predictive performance using a bootstrapping procedure.

**Results** Of the several algorithms developed using these data, predictive performance was shown to be better when VAS pain was included as a predictor and when the multivariate ordered probit regression method was used, rather than the beta regression method. The algorithms developed were shown to be comparable, in terms of predictive performance, to existing mapping studies despite the small sample size of the estimation data.

**Conclusion** It is hoped that the availability of these algorithms will facilitate the development of cost-effectiveness studies evaluating RA treatments in the Chinese health care setting.

**Keywords** Mapping · Rheumatoid arthritis · Cost effectiveness · Health utilities

## Introduction

Rheumatoid arthritis (RA) is a chronic, systemic inflammatory disorder and is primarily associated with progressive joint destruction and accompanied by pain, stiffness, and fatigue. If left untreated over the course of 10–20 years, RA may lead to significant disabilities and a severe reduction

in the patient's quality of life. The overall prevalence of RA is relatively constant across nations at 0.5–1% [1]. In China, RA is among the top 10 chronic diseases in terms of prevalence [2] and the second most common cause of disability [3]. Despite this heavy disease burden, the health care system in China does not currently cover biologic disease-modifying anti-rheumatic drugs (bDMARDs), which are an innovative class of therapeutic treatments for RA. This could potentially change however, as the reimbursement system in China looks set to move towards a value-based approach to health care coverage. In October 2010, the Ministry of Health (MoH) in China signed a memorandum of understanding with the National Institute for Health and Care Excellence (NICE) with a view to developing a new institution to promote quality and efficiency in health care. Furthermore, in an effort to facilitate future economic evaluations, Guidelines for Pharmacoeconomic Evaluations in China were published in 2011 [4]. These guidelines indicated a preference for the development of cost-effectiveness analyses (CEA) with quality-adjusted life years (QALY)

---

**Electronic supplementary material** The online version of this article (<https://doi.org/10.1007/s11136-018-1925-1>) contains supplementary material, which is available to authorized users.

---

✉ Thomas Patton  
thomas.patton@york.ac.uk

<sup>1</sup> Centre for Health Economics, University of York, York YO10 5DD, UK

<sup>2</sup> School of Biomedical Sciences and Pharmacy, University of Newcastle, Callaghan, NSW, Australia

<sup>3</sup> Department of Rheumatology, Taizhou No. 4 Renmin Hospital, Taizhou, Jiangsu, China

using health utility data derived from preference-based, generic health-related quality of life (HRQL) instruments.

Although there have been two studies to date conducting CEA for RA treatments in China [5, 6], neither of these studies used health utility evidence that was relevant to the Chinese setting. It is not uncommon to find that clinical trials collect disease-specific (non-preference based) HRQL instruments, as opposed to generic preference-based instruments, capable of deriving health utility values directly. For other countries, such as the UK and Spain, studies are available which map between disease-specific instruments, e.g. the Health Assessment Questionnaire (HAQ) for RA [7], and generic preference-based measures of HRQoL, such as the EQ-5D [8], via statistical modelling [9–19]. Although this approach has its shortcomings, in the absence of directly relevant health utility data, it is often considered acceptable by the national reimbursement jurisdictions, for example NICE [20]. One of the existing CEA studies on RA treatments in China used a mapping algorithm but the values correspond to a tariff from another country [5]. Ideally, health utility values employed in CEA studies should reflect the preferences of the jurisdiction under investigation given that there are important differences between tariffs from different countries due to cultural differences [21], and this can have a substantial impact on the findings of a CEA study [22]. To the authors' knowledge, there has yet to be any research conducted in a Chinese patient population to develop a mapping between HAQ and EQ-5D. In order to address the current deficiency of health utility evidence relevant for economic evaluation in the Chinese setting, the purpose of this study was to develop mapping algorithms linking the HAQ and EQ-5D-5L in a Chinese patient population. The authors anticipate that the availability of these algorithms will enable researchers to make use of the growing number of studies that have collected HAQ data in RA patients in China for the purposes of economic evaluation [23–26].

## Methods

### Data collection

A cross-sectional study was conducted in two tertiary referral hospitals, one in the Taizhou 5th Renmin Hospital in Jiangsu province and the other in Jianping County Hospital in Liaoning Province. The collection of data in separate regions of China was motivated by efforts to account for the significant heterogeneity, both in terms of economic development and living conditions, that exists across the Chinese population [27]. The study was approved by the Institutional Review Board and patients with RA were recruited, with informed consent, as a consecutive sample (i.e. data collection in all eligible patients until the desired sample size has

been achieved) between May and December 2013. Patients were eligible if they satisfied the following inclusion criteria:

- they have been previously documented as experiencing RA;
- they were not affected by any other type of musculoskeletal problem;
- they were aged 18 years or older, and identified as being capable of completing self-reported HRQoL questionnaires in Chinese;
- they had no serious psychiatric disorder or cognitive dysfunction; and
- they were not a current substance abuser.

Eligible subjects were asked to complete HAQ and EQ-5D questionnaires by a trained interviewer, in addition to a series of questions pertaining to their socio-demographic characteristics. Symptom severity and functional limitations were measured in subjects by physicians using the ACR classification of Global Functional Status. The interview procedure—including the interviewer asking the questions—was identical in the two recruiting centres.

## Instruments

### HAQ

The Health Assessment Questionnaire (HAQ), developed by Fries and colleagues in 1978 [28], has been widely administered and validated in patients for a range of rheumatic diseases. Moreover, it has become a de facto mandated outcome measure for clinical trials in rheumatoid arthritis as a component of the American College of Rheumatology criteria [29], which has subsequently been adopted for use in many RA studies in China [23–26, 30]. There were two components to the HAQ in this study: the HAQ Disability Index (HAQ-DI) and the HAQ pain visual analogue scale (VAS). The HAQ-DI consists of 20 items covering 8 domains assessing physical disabilities: dressing and grooming, arising, eating, walking, hygiene, reach, grip, and common daily activities. An additional 13 questions were included to assess the use of assistive devices in patients and a further 8 questions to assess whether or not patients received help from another person. Responses to each of the HAQ-DI questions are graded as follows: without any difficulty (0); with some difficulty (1); with much difficulty (2); and unable to do (3). The highest score for any component question in a category determines the category score. Two composite scores can be calculated, one with and one without the aids/devices element [27]. The HAQ pain VAS score is measured on a horizontal line where each end represents opposite ends of a continuum that is standardized to 15 centimeters in length. It is labelled with “no pain” and a score of 0 at one end,

and “very severe pain” and a score of 100 at the other end. Patients are instructed to place a vertical mark on the line to indicate the severity of their pain. A score from 0 to 100 is obtained based on the location of the respondent’s mark.

### EQ-5D-5L

EQ-5D-5L is a generic, preference-based HRQoL instrument with five dimensions including morbidity, self-care, usual activities, pain/discomfort, and anxiety/depression [31]. Each dimension has five response levels (no problems, slight problems, moderate problems, severe problems, and unable to/extreme problems) expressed by a 1-digit number, which results in a 5-digit number describing the respondent’s health state. Health utility index values were derived using a tariff taken from a recently developed study capturing health preferences in a sample of individuals from the general population in China [32]. The EQ-5D-5L was selected for use rather than the original EQ-5D instrument (EQ-5D-3L) on the grounds that the EQ-5D-5L was developed in an effort to improve on the original instrument’s perceived lack of sensitivity and also to reduce ceiling effects [31, 33].

### Statistical methods

A variety of statistical methods have been proposed in the published literature for the development of mapping algorithms [34]. Much of the debate in this area has revolved around the development of statistical methods capable of handling the unique distributional features of health utility values [35]. Ordinary least squares (OLS) regression methods are, at least in theory, inappropriate on the grounds that they do not guarantee predicted values will lie within a plausible range [36–38]. One proposed method for circumventing this problem is to employ regression estimators based on features of the beta distribution, which assumes that the dependent variable is restricted to a range of values between 0 and 1 [39, 40]. Another method known as the adjusted-limited dependent variable mixture model (ALDVMM) was developed to capture specific features of the UK version of the EQ-5D-3L tariff [38].

An alternative approach to mapping is to analyse the health state descriptions directly, rather than the health utility values, with modelling techniques including multinomial logit [41], ordered logistic [42], and ordered probit [43]. Unfortunately, these techniques fail to account for the correlations occurring between the different dimensions, which can give rise to misleading predictions. More recently, a study by Conigliani and colleagues showed that this issue is avoided by using a multivariate ordered probit method [44].

Mapping algorithms are derived in this study using two of the aforementioned regression methods: the beta regression method and a multivariate ordered probit regression.

The ALDVMM approach is not employed in this study on the grounds that this method was developed for the purposes of analysing health utilities obtained using the UK version of the EQ-5D-3L tariff. Although there are additional analytical methods beyond those discussed in this paper, the methods selected have been identified in ongoing research as holding the most potential for the purposes of mapping [45]. All analyses were undertaken in R, in which the beta regression method was implemented the ‘betareg’ package [46] and the multivariate approach was implemented using the ‘mvord’ package [47]. The index values of the EQ-5D-5L, which ranged between  $-0.391$  and  $1$ , had to be rescaled given that the beta regression handles values lying between  $0$  and  $1$ . The following equation was used to rescale values:

$$\text{rescaled value} = (\text{original value} + 0.391)/1.391.$$

Moreover, values lying at the either end of the distribution needed to be adjusted given that the beta regression cannot accommodate values of  $0$  or  $1$ . As such,  $1e-12$  was added to values equal to  $0$  and subtracted from values equal to  $1$ .

### Choice of predictors

There are multiple ways in which the HAQ could be incorporated into a mapping algorithm. The simplest approach would be to regress the overall HAQ score onto the EQ-5D. Alternatively, one could either use the HAQ item responses (i.e. questionnaire responses) or the subdomain scores (e.g. dressing, arising). Although the latter methods offer greater detail in the description of HRQoL effects, they may not be convenient for practical purposes. For instance, suppose that one wanted to predict health utility values using aggregated HAQ scores from a published study; mapping algorithms derived using item responses or subdomain scores would be incompatible with this evidence. Furthermore, the authors also felt that the overall HAQ score would be a more appropriate predictor given the small sample size of the estimation dataset (i.e. fewer parameters requiring estimation). Supplementary analyses are conducted with VAS pain score as an additional predictor following recommendations from Madan and colleagues [48].

### Validation of mapping algorithms

The internal validity of the mapping algorithms is explored using bootstrapping procedures. First, a bootstrap sample is drawn from the original estimation dataset, with replacement and a sample size equal to that of the original dataset. Mapping algorithms are then derived from the bootstrap sample, for each of the proposed modelling approaches. The estimated algorithms are then used to predict health utility values in the original estimation dataset. Finally, the

resulting predictions are compared to the observed health utility values using the root mean squared error (RMSE) statistic term. This process is repeated until 500 bootstrap iterations have been run.

When making predictions using the MV Probit approach, probabilities are required for each of the relevant health state descriptions. In theory, this would require predictions for each of the 3125 feasible states ( $5^5$ ) that can be constructed using the health state description of the EQ-5D-5L given that the MV Probit model accounts for correlations between each of the five dimensions. Thankfully, this is not necessary given that full health is the only state where between-dimension correlations have an impact on the index values derived using the Chinese EQ-5D-5L tariff. Consequently, the remaining probabilities reflect responses to specific dimensions, independent of the other dimensions. Once the probabilities have been predicted, EQ-5D values can then be scored. The authors decided to use the ‘expected utility’ method proposed in a study by Le and Doctor given that this ensures an exact calculation via an algebraic equation [49]. The following equation was used to conduct these calculations:

$$EQ5D\ prediction = (Prob_{Full\ Health}) \times 1 + (1 - Prob_{Full\ Health}) \times (1 - dis)\ dis = \{(Prob_{AD,2} \times 0.258) + (Prob_{MB,2} \times 0.345) + (Prob_{PD,2} \times 0.302) + (Prob_{SC,2} \times 0.253) + (Prob_{UA,2} \times 0.233)\} \times 0.191 + \{(Prob_{AD,3} \times 0.258) + (Prob_{MB,3} \times 0.345) + (Prob_{PD,3} \times 0.302) + (Prob_{SC,3} \times 0.253) + (Prob_{UA,3} \times 0.233)\} \times 0.458 + \{(Prob_{AD,4} \times 0.258) + (Prob_{MB,4} \times 0.345) + (Prob_{PD,4} \times 0.302) + (Prob_{SC,4} \times 0.253) + (Prob_{UA,4} \times 0.233)\} \times 0.832 + \{(Prob_{AD,5} \times 0.258) + (Prob_{MB,5} \times 0.345) + (Prob_{PD,5} \times 0.302) + (Prob_{SC,4} \times 0.253) + (Prob_{UA,4} \times 0.233)\}$$

where  $Prob_{FullHealth}$  represents the probability of a full health response and  $Prob_{i,j}$  represents the probability of a response of  $j$  on dimension  $i$ .

## Results

### Descriptive statistics and missing data

Table 1 provides descriptive statistics for the estimation sample used to derive mapping algorithms. The EQ-5D-5L responses in the complete case sample, without VAS pain scores ( $n = 130$ ), covered 43 of the 3,125 health states. The same finding was observed in the complete case sample with VAS pain scores ( $n = 125$ ). The proportion of patients reporting no health problems on any of the dimensions of the EQ-5D-5L was below 2% in the sample without VAS pain scores (2/130) and less than 1% in the sample with VAS pain scores (1/125).

Overall, the degree of missing data in the predictor variables was small, thus reducing the concerns one might have about obtaining less precise parameter estimates in a complete case analysis (i.e. due to reduced statistical power). In order to examine the validity of assuming that the data are missing completely at random (MCAR), logistic regressions were performed to explore any associations between the probability of a given predictor being missing and the values observed in the other predictors. No statistically significant associations were observed. This approach follows recommendations relating to the methods for best practice in the development of prediction models [50]. Additional tests found no statistically significant associations between the probability of a given predictor being missing and the values observed in other auxiliary variables (health care setting and region). In view of these findings, the authors decided that the risks posed by the missing data in the estimation sample were minimal and that a complete case analysis would be reasonable.

### Model estimates

Parameter estimates, alongside associated standard errors, for each of the mapping algorithms can be found in the supplementary materials. R scripts can also be found in the supplementary materials to implement the algorithms in mapping applications. This includes Cholesky decomposition matrices corresponding to each of the models. Models 1a and 2a refer to algorithms developed with HAQ score as the only covariate,

**Table 1** Descriptive statistics for the estimation sample

	Original data ( $n = 133$ )		Complete sample, without VAS pain ( $n = 130$ )	Complete sample, with VAS pain ( $n = 125$ )
	Mean (SD)	Missing data (%)	Mean (SD)	Mean (SD)
EQ-5D	0.66 (0.38)	1/133 (> 1%)	0.67 (0.38)	0.67 (0.37)
HAQ	1.22 (0.86)	2/133 (> 2%)	1.20 (0.86)	1.23 (0.84)
VAS pain	50.3 (24.6)	5/133 (> 4%)	–	50.0 (24.8)

while models 1b and 2b refer to the algorithms developed with both HAQ score and pain covariates. Note that predictions derived using the beta regression models need to be converted back onto the original scale between  $-0.391$  and  $1$  using the following equation:

$$\text{original scale} = (\text{predictions} \times 1.391) - 0.391.$$

### Validation of mapping algorithms

Table 2 contains the results from the bootstrapping procedures used to test the internal validity of the mapping algorithms. The MV Probit approach exhibited a lower RMSE in the majority of the bootstrap samples when compared to the beta regression approach. We can, therefore, conclude that the MV Probit modelling approach has a stronger predictive performance for the estimation sample under investigation when compared to the beta regression approach. The results also demonstrate that predictive performance is improved when the VAS pain covariate is included. Table 2 also shows that predictive performance is better in HAQ score less than 1 compared to those above. This finding is consistent with the broader mapping literature [35], i.e. that predictive performance is worse in poorer health states.

### Applying the preferred mapping algorithm using R code

The results of the bootstrapping exercise showed that the optimal mapping algorithm is model 1b, i.e. MV Probit method with both HAQ score and pain covariates, due to it having the best predictive performance of the four alternatives. There are several stages involved in the prediction of EQ-5D values using this algorithm and these have been set

out in an R script, along with a corresponding example, in the supplementary materials (see ‘Supplementary-Materials-2.docx’). This code has been annotated to demonstrate how the algorithm can be implemented in the R software package and how it can be adapted for the purposes of mapping applications involving data containing HAQ and VAS pain measures.

In the first stage of the code, the user is asked to specify whether the predicted values will be used to obtain deterministic or probabilistic cost-effectiveness results. In the case of the former, point estimates associated with each of the mapping parameters (i.e. regression coefficients, threshold values, and the error structure) will be used in the subsequent stages. Alternatively, if inputs for a probabilistic cost-effectiveness analysis are required, predictions will be made using parameter estimates that are randomly selected from a distribution that reflects the sample uncertainty surrounding the point estimates of each of the mapping parameters. In addition, the user is asked to load HAQ and VAS pain data to predict EQ-5D values (example data have been provided for the purposes of demonstration).

Stages 2 and 3 of the R script load the parameter estimates associated with model 1b, along with the corresponding variance–covariance matrix, which is used to construct sample distributions for each of the parameters. In stage 4, probabilities associated with different responses to items of the EQ-5D-5L are predicted. The code in stage 4 is based on the `get.prob()` function from the *mvord* package and has been adapted to accommodate external data. The final stage of the R script calculates the expected EQ-5D values by combining the probabilities derived in the previous step with the weights from the tariff estimated by Luo, Liu, and colleagues [32].

**Table 2** Internal validation of mapping algorithms

	All patients	Patients with HAQ less than 1	Patients with HAQ between 1 and 2	Patients with HAQ greater than 2
Models with HAQ covariate only				
RMSE (SD)				
MV probit	0.196 (0.013)	0.065 (0.003)	0.300 (0.019)	0.293 (0.039)
Beta	0.246 (0.021)	0.094 (0.018)	0.376 (0.037)	0.343 (0.011)
Samples where RMSE was lower in the MV probit (%)	100	99.6	99.8	91.8
Models with HAQ and VAS pain covariates				
RMSE (SD)				
MV probit	0.177 (0.022)	0.088 (0.022)	0.226 (0.027)	0.277 (0.031)
Beta	0.196 (0.016)	0.099 (0.014)	0.261 (0.028)	0.281 (0.014)
Samples where RMSE was lower in the MV probit (%)	94.2	90.8	95.2	63.8

## Discussion

This study is the first to develop mapping algorithms linking HAQ scores to a preference-based measure of HRQoL in a Chinese patient population. Until now, the absence of such evidence has been a major obstacle for researchers seeking to find relevant evidence to populate CEAs comparing RA treatments in China. The availability of the algorithms in this study is timely given that there has been an increasing number of CEAs conducted in the Chinese health care setting [51]. Moreover, recent actions by the Ministry of Health in China point towards an increased use of HTA for policy-making purposes [52, 53]. It is hoped that the availability of R scripts for the implementation of the algorithms will facilitate their usage in applied CEA studies.

The mapping algorithms were derived using a cross-sectional dataset conducted in two hospitals in China that collected EQ-5D-5L, HAQ scores, and VAS pain scores in patients with RA. Two established statistical methodologies—the beta regression and the MV ordered probit regression—were employed to develop mapping algorithms. These fundamentally different approaches to mapping were selected in light of ongoing research identifying them as holding the most potential for overcoming the well-established shortcomings associated with standard OLS regression methods for the purposes of mapping.

The predictive performance of the mapping algorithms developed in this study was tested by estimating the RMSE for each model specification using a bootstrapping procedure. Using this approach, the MV ordered probit model exhibited lower prediction errors when compared to the beta regression model. Prediction errors were also lower in those models including a VAS pain covariate. Overall, the predictive performance of the MV ordered probit models was consistent with the range of RMSEs (0.1644–0.207) observed in mapping studies for RA in the published literature [11, 16, 18, 19, 38, 43]. To the authors' knowledge, this study is the first to compare the predictive performance of the beta and MV ordered probit methods for the purposes of mapping in cost-effectiveness applications. In this study, the computational burden incurred when predicting EQ-5D values using the MV ordered probit method was not as extensive as it could potentially have been. This was because, aside from the full health state, between-dimension correlations did not have an impact on the resulting values obtained when applying the Chinese EQ-5D-5L tariff. It is important to acknowledge that a different approach may have been needed for a tariff with a different model specification.

It is important to recognize that there are several limitations associated with the evidence used in this study. The sample size of the dataset used to derive the mapping algorithms was small, with only 130 patients (125 in models

including the VAS pain as a predictor), in comparison to similar studies in the published literature [9–19]. A larger sample would be preferable given that an increase in statistical power leads to improved precision in the estimation of parameters. Ultimately, this is important in the context of a decision model as it can potentially result in reduced parameter uncertainty in the cost-effectiveness estimates. Although the bootstrapping exercise in this study provides an assessment of the internal validity of the mapping procedures considered in this study, it does not tell us anything about the generalizability (or external validity) of the algorithms. Ideally, researchers should investigate the generalizability of prediction models using an independent dataset [50]. Unfortunately, the lack of external data meant that the authors were unable to do this.

Another potential limitation in this study is the exclusion of patients with other musculoskeletal problems, which may confound the relationship between HAQ scores and EQ-5D values. The degree to which this may ultimately affect cost-effectiveness results for a given research question will depend on the prevalence of other musculoskeletal comorbidities in the patient population under investigation. Researchers should yield caution when applying the mapping algorithms from this paper if the prevalence of such comorbidities is high.

There are various ways in which the HAQ data could have been specified in terms of predictors included in the mapping algorithms. Some previous studies have captured HAQ data in the form of categorical responses to the questionnaire items using dummy variables [11, 13, 16]. Backwards or forwards stepwise selection procedures are typically used to identify significant items rather than using all 42 individual items. Another approach has been to specify predictors reflecting the 8 domains within the HAQ measure [11], i.e. dressing/grooming, rising, eating, walking, hygiene, reach, grip, activities. One advantage of specifying HAQ data in terms of item-level responses or domain-level scores, rather than overall index scores, is that they have the potential to account for a higher proportion of the variance in the dependent variable. However, mapping algorithms containing item-level responses as predictors can only be used to predict EQ-5D values in conjunction with HAQ data at the individual patient level; consequently, this rules out the possibility of using evidence from the published literature. Moreover, the specification of HAQ data in terms of item-level responses or domain-level scores implies a larger number of predictors and, consequently, a reduction in the statistical power for a given sample size. The overall HAQ score was considered to be the most appropriate predictor for the purposes of this study given the small sample size of the estimation dataset (i.e. fewer parameters requiring estimation).

## Conclusions

In recent years, the availability of mapping algorithms linking HAQ to generic measures of HRQL has facilitated the development of cost-effectiveness studies evaluating treatments for RA [20]. This study is the first, to the authors' knowledge, to have developed a mapping algorithm between HAQ and EQ-5D in a Chinese patient population. The estimation sample was obtained from a cross-sectional study that collected data in RA patients in two tertiary referral hospitals in China. Of the several algorithms developed using these data, predictive performance was shown to be better when VAS pain was included as a predictor and when the multivariate ordered probit regression method was used, rather than the beta regression method. The algorithms developed were shown to be comparable, in terms of predictive performance, to existing mapping studies despite the small sample size of the estimation data. It is hoped that the availability of these algorithms will facilitate the development of cost-effectiveness studies evaluating RA treatments in the Chinese health care setting.

**Funding** Aside from the data collection that was financed by the Chinese Scholarship Council, there was no additional funding supporting this work.

## Compliance with ethical standards

**Conflict of interest** The authors can confirm that there are no conflicts of interest to declare.

**Ethical approval** All the authors confirm that the research was approved by the Human Research Ethics Administration at the University of Newcastle and have therefore been performed in accordance with ethical standards laid down in the 1964 Declaration of Helsinki.

**Informed consent** Informed consent was obtained from all individual participants included in the study.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## References

- Silman, A. J., & Hochberg, M. C. (2001). *Epidemiology of the rheumatic diseases* (2nd edn.). Oxford: Oxford University Press.
- China Health Statistical Yearbook 2012. 2012. Chinese Center for Disease Control and Prevention: Beijing.
- Documentation of the Second China National Samples Survey on Disability. 2007, China Statistics Press.
- International Society for Pharmacoeconomics and Outcomes Research. Pharmacoeconomic guidelines around the world: China. 2011. Retrieved March 12, 2017, from <https://www.ispor.org/PEguidelines/COUNTRYSPECIFIC.asp>.
- Wu, B., Wilson, A., Wang, F. F., et al. (2012). Cost effectiveness of different treatment strategies in the treatment of patients with moderate to severe rheumatoid arthritis in china. *PLoS ONE*, 7(10), e47373.
- Zhang, Y., Yang, L., Dong, P., Xie, X., & Tang, B. (2012). PMS14 cost-effectiveness analysis of etanercept in the treatment of rheumatoid arthritis in China. *Value in Health*, 15(7), A673.
- Fries, J. F., Spitz, P., Kraines, R. G., & Holman, H. R. (1980). Measurement of patient outcome in arthritis. *Arthritis & Rheumatology*, 23(2), 137–145.
- The EuroQol Group. (1990). EuroQol—A new facility for the measurement of health-related quality of life. *Health Policy*, 16(3), 199–208.
- Ariza-Ariza, R., Hernández-Cruz, B., Carmona, L., et al. (2006). Assessing utility values in rheumatoid arthritis: A comparison between time trade-off and the EuroQol. *Arthritis Care & Research*, 55(5), 751–756.
- Witney, A. G., Treharne, G. J., Tavakoli, M., et al. (2006). The relationship of medical, demographic and psychosocial factors to direct and indirect health utility instruments in rheumatoid arthritis. *Rheumatology*, 45(8), 975–981.
- Bansback, N., Marra, C., Tsuchiya, A., et al. (2007). Using the health assessment questionnaire to estimate preference-based single indices in patients with rheumatoid arthritis. *Arthritis Care & Research*, 57(6), 963–971.
- Lillegraven, S., Kristiansen, I. S., & Kvien, T. K. Comparison of utility measures and their relationship with other health status measures in 1041 patients with rheumatoid arthritis. *Annals of the Rheumatic Diseases*. 2010 69:1762–1767
- Wolfe, F., Michaud, K., & Wallenstein, G. (2010). Scale characteristics and mapping accuracy of the US EQ-5D, UK EQ-5D, and SF-6D in patients with rheumatoid arthritis. *The Journal of Rheumatology*, 37(8), 1615–1625.
- Carreño, A., Fernández, I., Badia, X., et al. (2011). Using HAQ-DI to estimate HUI-3 and EQ-5D utility values for patients with rheumatoid arthritis in Spain. *Value in Health*, 14(1), 192–200.
- Alava, M. H., Wailoo, A., Wolfe, F., & Michaud, K. (2013). The relationship between EQ-5D, HAQ and pain in patients with rheumatoid arthritis. *Rheumatology*, 52(5), 944–950.
- Kontodimopoulos, N., Bozios, P., Yfantopoulos, J., & Niakas, D. (2013). Longitudinal predictive ability of mapping models: Examining post-intervention EQ-5D utilities derived from baseline MHAQ data in rheumatoid arthritis patients. *The European Journal of Health Economics*, 14(2), 307–314.
- Hoshi, D., Tanaka, E., Igarashi, A., et al. (2016). Profiles of EQ-5D utility scores in the daily practice of Japanese patients with rheumatoid arthritis; Analysis of the IORRA database. *Modern Rheumatology*, 26(1), 40–45.
- Kim, H. L., Kim, D., Jang, E. J., et al. (2016). Mapping health assessment questionnaire disability index (HAQ-DI) score, pain visual analog scale (VAS), and disease activity score in 28 joints (DAS28) onto the EuroQol-5D (EQ-5D) utility score with the KOREan Observational study Network for Arthritis (KORONA) registry data. *Rheumatology International*, 36(4), 505–513.
- Nair, S. C., Welsing, P. M., Marijnissen, A. K., et al. (2015). Does disease activity add to functional disability in estimation of utility for rheumatoid arthritis patients on biologic treatment? *Rheumatology*, 55, 94.
- Malotki, K., Barton, P., Tsourapas, A., et al. (2011). Adalimumab, etanercept, infliximab, rituximab and abatacept for the treatment of rheumatoid arthritis after the failure of a tumour necrosis factor inhibitor: A systematic review and economic evaluation. *Health Technology Assessment*, 15(14).

21. Norman, R., Cronin, P., Viney, R., et al. (2009). International comparisons in valuing EQ-5D health states: A review and analysis. *Value in Health*, *12*(8), 1194–1200.
22. Pennington, B., & Davis, S. (2014). Mapping from the Health Assessment Questionnaire to the EQ-5D: The impact of different algorithms on cost-effectiveness results. *Value in Health*, *17*(8), 762–771.
23. An, Y., Liu, T., He, D., Wu, L., Li, J., Liu, Y., et al. (2017). The usage of biological DMARDs and clinical remission of rheumatoid arthritis in China: A real-world large scale study. *Clinical Rheumatology*, *36*(1), 35–43.
24. Li, Z., Zhang, F., Kay, J., Fei, K., Han, C., Zhuang, Y., Wu, Z., & Hsia, E. C. (2016). Efficacy and safety results from a Phase 3, randomized, placebo-controlled trial of subcutaneous golimumab in Chinese patients with active rheumatoid arthritis despite methotrexate therapy. *International Journal of Rheumatic Diseases*, *19*(11), 1143–1156.
25. Kremer, J. M., Blanco, R., Brzosko, M., Burgos-Vargas, R., Halland, A. M., Vernon, E., Ambs, P., & Fleischmann, R. (2011). Tocilizumab inhibits structural joint damage in rheumatoid arthritis patients with inadequate responses to methotrexate: Results from the double-blind treatment phase of a randomized placebo-controlled trial of tocilizumab safety and prevention of structural joint damage at one year. *Arthritis & Rheumatology*, *63*(3), 609–621.
26. US National Library of Medicine. *Real-World Outcome of Adalimumab on Rheumatoid Arthritis Patients in China*. Retrieved February 18, 2018, from <https://clinicaltrials.gov/ct2/show/NCT02668640>.
27. Hu, H., Luan, L., Yang, K., & Li, S. C. (2016). Psychometric validation of Chinese Health Assessment Questionnaire for use in rheumatoid arthritis patients in China. *International Journal of Rheumatic Diseases*, *20*, 1987–1992.
28. Fries, J. F., Spitz, P. W., & Young, D. Y. (1981). The dimensions of health outcomes: the health assessment questionnaire, disability and pain scales. *The Journal of Rheumatology*, *9*(5), 789–793.
29. Hochberg, M. C., Chang, R. W., Dwosh, I., Lindsey, S., Pincus, T., & Wolfe, F. (1992). The American College of Rheumatology 1991 revised criteria for the classification of global functional status in rheumatoid arthritis. *Arthritis & Rheumatism*, *35*(5)(1), 498–502.
30. Wang, G., Mu, R., & Xu, H. (2015). Management of rheumatoid arthritis in People's Republic of China—Focus on tocilizumab and patient considerations. *International Journal of General Medicine*, *8*, 187–194.
31. Herdman, M., Gudex, C., Lloyd, A., et al. (2011). Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Quality of Life Research*, *20*(10), 1727–1736.
32. Luo, N., Liu, G., Li, M., et al. (2017). Estimating an EQ-5D-5L value set for China. *Value in Health*, *20*, 662–669.
33. Janssen, M. F., Bonsel, G. J., & Luo, N. (2018). Is EQ-5D-5L better than EQ-5D-3L? A head-to-head comparison of descriptive systems and value sets from seven countries. *Pharmacoeconomics*, *36*, 675–697.
34. Dakin, H. (2013). Review of studies mapping from quality of life or clinical measures to EQ-5D: An online database. *Health and Quality of Life Outcomes*, *11*(1), 151.
35. Longworth, L., & Rowen, D. (2011). The use of mapping methods to estimate health state utility values. NICE DSU Technical Support Document 10.
36. Austin, P. C. (2002). A comparison of methods for analyzing health-related quality-of-life measures. *Value in Health*, *5*(4), 329–337.
37. Chuang, L. H., & Kind, P. (2009). Converting the SF-12 into the EQ-5D. *Pharmacoeconomics*, *27*(6), 491–505.
38. Alava, M. H., Wailoo, A. J., & Ara, R. (2012). Tails from the peak district: adjusted limited dependent variable mixture models of EQ-5D questionnaire health state utility values. *Value in Health*, *15*(3), 550–561.
39. Basu, A., & Manca, A. (2012). Regression estimators for generic health-related quality of life and quality-adjusted life years. *Medical Decision Making*, *32*(1), 56–69.
40. Hunger, M., Döring, A., & Holle, R. (2012). Longitudinal beta regression models for analyzing health-related quality of life scores over time. *BMC Medical Research Methodology*, *12*(1), 144.
41. Gray, A. M., Rivero-Arias, O., & Clarke, P. M. (2006). Estimating the association between SF-12 responses and EQ-5D utility values by response mapping. *Medical Decision Making*, *26*(1), 18–29.
42. Van Hout, B., Janssen, M. F., Feng, Y. S., et al. (2012). Interim scoring for the EQ-5D-5L: mapping the EQ-5D-5L to EQ-5D-3L value sets. *Value in Health*, *15*(5), 708–715.
43. Hernández Alava, M., Wailoo, A., Wolfe, F., & Michaud, K. (2014). A comparison of direct and indirect methods for the estimation of health utilities from clinical outcomes. *Medical Decision Making*, *34*(7), 919–930.
44. Conigliani, C., Manca, A., & Tancredi, A. (2015). Prediction of patient-reported outcome measures via multivariate ordered probit models. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *178*(3), 567–591.
45. Research Councils, U. K. *Modelling generic preference based outcome measures—Development and comparison of methods*. Retrieved June 26, 2017, from <http://gr.rcuk.ac.uk/project/5A46044C-DC5B-4A0D-B21E-863FCFF87114>.
46. Zeileis, A., Cribari-Neto, F., Gruen, B., et al. Package ‘betareg’.
47. Hirk, R., Hornik, K., & Vana, L. (2017). *mvord: An R Package for Fitting Multivariate Ordinal Regression Models*.
48. Madan, J., Ades, T., Barton, P., et al. (2015). Consensus decision models for biologics in rheumatoid and psoriatic arthritis: Recommendations of a multidisciplinary working party. *Rheumatology and Therapy*, *2*(2), 113–125.
49. Le, Q. A., & Doctor, J. N. (2011). Probabilistic mapping of descriptive health status responses onto health state utilities using Bayesian networks: an empirical analysis converting SF-12 into EQ-5D utility index in a national US sample. *Medical Care*, *49*, 451–460.
50. Steyerberg, E. (2008). *Clinical prediction models: a practical approach to development, validation, and updating*. New York: Springer.
51. Thorat, T., Lin, P. J., & Neumann, P. J. (2015). The state of cost-utility analyses in Asia: A systematic review. *Value in Health Regional Issues*, *6*, 7–13.
52. Kennedy-Martin, T., Mitchell, B. D., Boye, K. S., et al. (2014). The health technology assessment environment in mainland China, Japan, South Korea, and Taiwan—Implications for the evaluation of diabetes mellitus therapies. *Value in Health Regional Issues*, *3*, 108–116.
53. Chen, Y., Banta, D., & Tang, Z. (2009). Health technology assessment development in China. *International Journal of Technology Assessment in Health Care*, *25*(S1), 202–209.