

ORIGINAL ARTICLE

Mapping of the World Health Organization's Disability Assessment Schedule 2.0 to disability weights using the Multi-Country Survey Study on Health and Responsiveness

Joran Lokkerbol^{1,2} | Ben F. M. Wijnen^{1,3}  | Somnath Chatterji⁴  |
Ronald C. Kessler²  | Dan Chisholm⁵

¹Center of Economic Evaluation & Machine Learning, Trimbos Institute (Netherlands Institute of Mental Health and Addiction), Utrecht, The Netherlands

²Department of Health Care Policy, Harvard Medical School, Boston, Massachusetts, USA

³Department of Clinical Epidemiology and Medical Technology Assessment, Maastricht University Medical Center, Maastricht, The Netherlands

⁴Department of Data and Analytics, World Health Organization, Geneva, Switzerland

⁵Division of Country Health Policies and Systems, Regional Office for Europe, World Health Organization, Copenhagen, Denmark

Correspondence

Joran Lokkerbol, Center of Economic Evaluation & Machine Learning, Trimbos Institute (Netherlands Institute of Mental Health and Addiction), Da Costakade 45 3521 VS Utrecht, Netherlands.
Email: jlokkerbol@trimbos.nl

Funding information

Netherlands Organisation for Health Research and Development (ZonMw) Mental Healthcare Fellowship, Grant/Award Number: # 636320003; World Health Organization

Abstract

Objectives: To develop and test an internationally applicable mapping function for converting WHODAS-2.0 scores to disability weights, thereby enabling WHODAS-2.0 to be used in cost-utility analyses and sectoral decision-making.

Methods: Data from 14 countries were used from the WHO Multi-Country Survey Study on Health and Responsiveness, administered among nationally representative samples of respondents aged 18+ years who were non-institutionalized and living in private households. For the combined total of 92,006 respondents, available WHODAS-2.0 items (for both 36-item and 12-item versions) were mapped onto disability weight estimates using a machine learning approach, whereby data were split into separate training and test sets; cross-validation was used to compare the performance of different regression and penalized regression models. Sensitivity analyses considered different imputation strategies and compared overall model performance with that of country-specific models.

Results: Mapping functions converted WHODAS-2.0 scores into disability weights; R-squared values of 0.700–0.754 were obtained for the test data set. Penalized regression models reached comparable performance to standard regression models but with fewer predictors. Imputation had little impact on model performance. Model performance of the generic model on country-specific test sets was comparable to model performance of country-specific models.

Conclusions: Disability weights can be generated with good accuracy using WHO-DAS 2.0 scores, including in national settings where health state valuations are not directly available, which signifies the utility of WHODAS as an outcome measure in evaluative studies that express intervention benefits in terms of QALYs gained.

KEYWORDS

disability weight, mapping function, multi-country survey study on health and responsiveness, WHODAS-2.0

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2021 The Authors. International Journal of Methods in Psychiatric Research published by John Wiley & Sons Ltd.

1 | INTRODUCTION

From the public health perspective, a key challenge for policy and practice is to promote and provide a healthcare system that is effective, acceptable, and sustainable (Donahue et al., 2018; James et al., 2018; Schmidt et al., 2015). An increasing interest and investment in effectiveness research has accompanied the pursuit of these goals (van Velden et al., 2005; Williams et al., 2008). One of the enduring difficulties in using effectiveness research for contributing to public health goals, is the lack of comparability of outcomes used in different intervention studies (Afzali et al., 2013; Walker et al., 2010). Transdiagnostic measures of outcome such as functioning, as assessed in such scales as the Short Form-36 (SF-36) (Anderson et al., 1996) or World Health Organization Disability Assessment Schedule 2.0 (WHODAS 2.0), provide a degree of comparability across interventions to help support decision-making (Ustün et al., 2010). However, clinical studies often focus on disease-specific symptomatology or personal and social functioning (henceforth functioning) outcomes rather than generic measures such as Sort Form-6 (SF-6) or EuroQol-5D (Dolan, 1997), as the latter are not always seen as clinically relevant or sufficiently responsive to change [e.g. (Wijnen et al., 2018)]. In addition, for sectoral decision-making or priority-setting exercises that need to combine or compare changes in morbidity and/or mortality, studies principally report cost per Quality-Adjusted Life Year (QALY) or cost per Disability-Adjusted Life Year (DALY) (Drummond et al., 2015; Homedes, 1996). The DALY methodology was developed as a generic framework for quantifying disease burden, such that the burden of different health conditions can be directly compared and ranked using a single metric, thereby informing budget allocation decisions and research agenda setting (Salomon et al., 2015). Use of QALYs or DALYs as an outcome measure in sectoral cost-effectiveness analysis requires the estimation of utilities or disability weights, which in turn requires a population-based survey of health state preferences against which a profile score based on measures such as the SF-6 or EQ-5D can be mapped (Balestroni & Bertolotti, 2012; Brazier et al., 2002). Such studies are mainly available in high-income countries (Devlin & Brooks, 2017).

Population-based health state surveys have been carried out for WHODAS 2.0 in several regions of the world and this scale can be used to generate health state valuations for deriving comparable outcomes across different study populations. For example, health state valuations were derived using WHODAS 2.0 by Buttorff et al. (2012) in a cost-effectiveness study for common mental disorders in India (Buttorff et al., 2012). As WHODAS 2.0 data are available for multiple countries, a generic mapping from WHODAS 2.0 to disability weight could be derived across countries if the transformation rules were developed. This would allow the use of a more clinically relevant measure such as WHODAS 2.0, while also facilitating the policy making perspective by eliciting QALYs as an outcome in intervention studies. WHODAS 2.0 is applicable across all conditions, including mental, behavioural and neurological as well as other chronic conditions in both clinical and general population

settings across cultures and is sensitive to change (Ustün et al., 2010; Garin et al., 2010). WHODAS 2.0 captures functioning in six life domains: (1) Cognition: understanding and communicating; (2) Mobility: moving and getting around; (3) Self-care: attending to one's hygiene, dressing, eating and staying alone; (4) Getting along: interacting with other people; (5) Life activities: domestic responsibilities, leisure, work and school, and; (6) Participation: joining in community activities, participating in society.

Between 2000–2001, the Multi-Country Survey Study on Health and Responsiveness (MCSS) was conducted in 61 countries (Üstün et al., 2001). In 14 countries, a more extensive face-to-face household survey was carried out, that estimated disability weights for all respondents (estimated by a health state valuation function (Üstün et al., 2001)) and included a subset of the WHODAS 2.0. It is possible to use this dataset to develop an algorithm that maps WHODAS 2.0 scores into disability weights (Wijnen et al., 2018). Such a mapping would facilitate the assessment of burden of disease for conditions for which disability weights have not been directly assessed by using WHODAS 2.0 to generate disability weight measures when such measures were not obtained directly. Also, this mapping helps in generating cost-effectiveness outcomes in intervention studies. The aim of the current report is to present the result of an effort to develop a generic, country-independent, mapping algorithm converting WHODAS 2.0 items into disability weights using MCSS-data for all countries where the household mode survey was administered.

2 | METHODS

This study is reported using the standards of the Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD) statement (Collins et al., 2015). Direct response mapping was used to estimate disability weights based on WHODAS 2.0-36 items and WHODAS 2.0-12 items. In direct mapping, a mapping function (such as a regression equation) was used to predict disability weights using scores from the WHODAS 2.0 items as predictors. Such a mapping function can be applied to a new dataset to convert the source measure into the target measure based on the assumption that the associations of WHODAS 2.0 items with disability weights are generalizable (Wijnen et al., 2018).

2.1 | Data

The “household mode” of the MCSS was administered in nationally representative samples (male and female adults aged above 18 years, non-institutionalized and living in private households) in the following 14 countries: China, Colombia, Egypt, Georgia, Indonesia, India, Iran, Lebanon, Mexico, Nigeria, Singapore, Slovakia, Syria, and Turkey. These countries had a combined total of 92,006 observations, with sample sizes varying from $n = 1,183$ (Slovakia) to $n = 9,994$ (Indonesia) (Üstün et al., 2001). The MCSS-survey included only 19 of the 36 items that comprise the WHODAS 2.0 full version

and eight items of the WHODAS 2.0 short version since the MCSS health state description section had to include several additional items given the scope of that study (see appendix 1 for a list of the included items of the WHODAS 2.0 in the MCSS survey). Hence, the mapping was performed using both these 19 as well as these eight items as predictors for disability weights. Furthermore, the MCSS survey data included the demographic variables age, gender, marital status, educational level, and work status.

The MCSS-data were used to estimate disability weights using a health state valuation function based on a set of six core domain levels (i.e. mobility, self-care, usual activities, pain, affect, cognition), rated on a 1 to 5 Likert scale in which 1 indicated “No difficulty” and 5 indicated “Extreme difficulty/cannot do” (Murray & Evans, 2003a, 2003b). Respondents were provided descriptions of hypothetical health states along a set of core domains and asked to evaluate these health states using the visual analogue scale (VAS). Next, 18 different regression models that varied in such characteristics as the level of interactions considered were estimated in which VAS-scores were related to vignette-adjusted levels on six domains (Murray & Evans, 2003a, 2003b). Corresponding VAS-results were adjusted using a scale distortion parameter that was based on the multi-method exercises included in the MCSS (i.e. time trade-off, standard gamble, and person trade-off method) to adjust for end-aversion bias in the visual analogue scale (Murray & Evans, 2003a, 2003b). This resulted in the following transformation in which raw VAS-scores were transformed into adjusted VAS-scores, representing the health state valuation:

$$VAS_{adjusted} = \left(1 - (1 - VAS_{raw}/100)^{\left(\frac{1}{0.64}\right)} \right) * 100,$$

which was taken from Murray & Evans, (2003a, 2003b) and where the constant 0.64 was determined to be the optimal value for the scale distortion parameter.

Out of the 18 available regression models estimating raw VAS-scores based on the six core domains reported in (Murray & Evans, 2003a, 2003b), we applied the main effects model based on the assumption of a normal distribution for reasons of transparency and interpretability. This decision was justified based on evidence that there was only minimal impact of adding interaction terms to the health state valuation function (Murray & Evans, 2003a, 2003b).

Moreover, regression models with interactions contained large coefficients with wide confidence intervals that might have been due to overfitting (Sayak, 2018). Lastly, the model assuming the normal distribution performed better in the mid-range of observed VAS values (Murray & Evans, 2003a, 2003b), which is the range most relevant to the clinical populations to which we intend the mapping functions to be applied. As the health state valuation function was based on six core domains (see above) we only included patients who completed all six questions, and hence, for whom it was possible to estimate a raw VAS-score. Adjusted VAS-scores, rescaled to disability weights, were then used as dependent outcome (see Figure 1).

2.2 | Analysis

The mapping algorithm was constructed using a machine learning approach (Wiens & Shenoy, 2018). Hence, the best performing mapping function/algorithm was constructed using the following sequential steps:

- 1) *Selecting predictors.* Models were run using information on WHODAS 2.0–36 and WHODAS 2.0–12 and/or demographics and/or individual countries as predictors, resulting in the following 10 sets of predictors for which the ability to predict disability weight was compared: (1) all available individual WHODAS 2.0–36 items; (2) all available individual WHODAS 2.0–36 items and demographics; (3) WHODAS 2.0–36 domain scores; (4) WHODAS 2.0–36 domain scores and demographics; (5) all available WHODAS 2.0–36 items with demographics and country dummies; (6) all available WHODAS 2.0–36 items with demographics and country dummies and country interactions (with all other variables). As the WHODAS 2.0–12 does not use domain scores, only the models considering individual items were included: (7) all available individual WHODAS 2.0–12 items; (8) all available individual WHODAS 2.0–12 items and demographics; (9) all available WHODAS 2.0–12 items with demographics and country dummies; (10) all available WHODAS 2.0–12 items with demographics and country dummies and country interactions (with all other variables).
- 2) *Splitting the data in a training and test set.* The dataset was split into a training set used for model selection and a test (hold-out) set

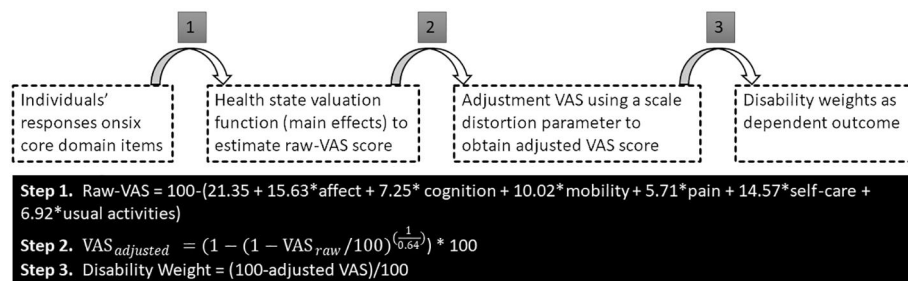


FIGURE 1 Schematic overview of the derivation of disability weight

used for model assessment by selecting a random 75% of the data for training and the remaining 25% of the data for testing for each country in the dataset.

- 3) *Data preparation.* Before model fitting, missing data were imputed. Missing demographic information was imputed using the median (for continuous variables) or a label “missing” (for categorical variables). In line with the Manual for WHO Disability Assessment Schedule, missing items were imputed using the mean of the other items within the same domain (Üstün et al., 2010). If not available, missing items were imputed using the participants' mean score on all available domains. If no domains were available for a participant, column means are used for imputation per missing item. (See below for information on frequency of missingness.). In the Singaporean questionnaire, age was asked in a categorical way. Hence, this age variable was converted into a numeric variable by imputing the median age for each category.
- 4) *Fitting various statistical learning models (i.e. mapping algorithms) on the training set.* To maximize interpretability, linear regression (i.e. ordinary least squares) and least absolute shrinkage and selection operator (LASSO) regression were used as statistical learning methods. LASSO augments the linear regression approach of minimizing the sum of squared errors with a penalty term (i.e. lambda or shrinkage parameter) proportional to the size of each (absolute) standardized beta, which has the effect of excluding variables from the model, and thus leading to a simpler model (James et al., 2013). The hyperparameter for the LASSO regressions, being the size of the penalty term, was tuned by means of a grid search going from practically no ($\lambda = 10^{-3}$) to a large penalty ($\lambda = 10^3$). To optimize the fit of the various models while preventing overfitting of the training set, 10-fold cross-validation was used. The root mean squared error (RMSE) and R-squared were considered for each model and RMSE was used to determine the best performing model. The RMSE represents the standard deviation of the prediction errors, such that the lower the RMSE, the better the model fit. The R-squared expresses how much variance is explained by the model relative to how much variance there is to explain (Field, 2013). The higher the R-squared, the better the model fit. Analyses were done in R (4.0.3), a statistical programming language (Chambers, 2008). The *caret* package was used for the machine learning analyses, including cross-validation and hyperparameter tuning (Kuhn, 2012).
- 5) *Evaluating the model on the test set.* For both WHODAS 2.0–36 and WHODAS 2.0–12, the models with the best cross-validated performance on the training set with and without country as predictor (i.e. to obtain both a generic and country-specific mapping algorithm) were assessed by evaluating the performance on the test set (after applying the same rule-based data preparation steps to the test set). Moreover, to provide an estimation on how well the model generalizes to other countries not included in this study, an analysis was performed in which a model was trained on 13 out of the 14 country-specific datasets with the

remaining 14th country-specific dataset serving as a test set. Model training on the 13 countries was done using leave-one-country-out cross-validation.

The syntax of the analyses is available upon request at the corresponding author.

2.3 | Sensitivity analyses

Model performance on the test set was determined using alternative strategies for handling missing data: (1) by considering only records with no missing data; (2) by imputing missing WHODAS domain scores using the mean of those domain scores for other respondents, instead of the mean of other domain scores within the same respondent; and; (3) by imputation using the k nearest neighbours (kNN) algorithm as implemented in the *caret* package (using the default of five nearest neighbours). kNN imputation imputes missing data using the (non-missing) values from the k closest neighbours to the observation with missing data. In addition, models were estimated for each country individually.

3 | RESULTS

3.1 | Sample characteristics

Individuals' responses on at least one core domain were missing in 3,772 (4.1%) respondents, resulting in a missing disability weight and hence exclusion from the study. On average, these missing respondents were older compared to completers (46.7 years old compared to 39.6 years old). The majority of the missing values occurred in the Colombian dataset (2,167; 57.4%). In the resulting dataset ($N = 88,234$), WHODAS subdomain scores were missing for “Understanding and communication” in 549 respondents (0.6%), “Getting around” in 143 (0.2%) respondents; “Self-care” in 4,234 (4.8%) respondents; “Getting along with people” in 23,765 (26.9%) respondents; “Life activities” in 30,831 (34.9%) respondents; and “Participation in society” in 5,235 (5.9%) respondents.

An overview of the sample characteristics by country is presented in Table 1. Except for Singapore, all demographic variables had less than 2% missing values. For Singapore, information regarding educational background was missing for 13% of participants. Mean age ranged from an average of 36.0 years for Nigeria to 45.6 years for Georgia. For all countries, most participants reported to be currently married. The percentage of females was 47.1% on average, ranging from 40.4% to 65.4% between individual countries. Educational background varied substantially between countries with some countries reporting around 50% or more to have followed less than primary school (e.g. Egypt, Indonesia, and India with a range of participants reporting less than primary school of 49.3% to 62.3%), versus 1.2% and 3.3% in Georgia and Turkey.

TABLE 1 Sample characteristics by country

Characteristic	China (N = 9486)	Colombia (N = 8158)	Egypt (N = 4490)	Georgia (N = 9847)	Indonesia (N = 9994)	India (N = 5144)	Iran (N = 9718)	Lebanon (N = 3246)	Mexico (N = 4813)	Nigeria (N = 5108)	Singapore (N = 6216)	Slovakia (N = 1183)	Syria (N = 9344)	Turkey (N = 5207)
Age, mean (SD)	39.7 (14.1)	40 (15.8)	39.1 (14.3)	45.6 (16.8)	39.9 (14.9)	40.1 (16.4)	37.6 (15.6)	42.2 (16.6)	41.8 (16.4)	36.0 (16)	41.0 (13.9)	42.5 (16.6)	37.7 (15)	33.4 (12.1)
Missing	113 (1.2%)	0 (0%)	6 (0.13%)	5 (0.1%)	196 (2.0%)	0 (0%)	40 (0.4%)	24 (0.8%)	1 (0.0%)	40 (0.8%)	0 (0%)	31 (2.7%)	5 (0.1%)	69 (1.4%)
Female, n (%)	4411 (46.8%)	3919 (65.4%)	2508 (56.1%)	5559 (57.7%)	5436 (54.8%)	2747 (53.4%)	4960 (52.2%)	1660 (51.9%)	1943 (40.4%)	3094 (61.4%)	3090 (49.7%)	633 (54.5%)	4556 (53.0%)	2184 (42.7%)
Missing	2 (0.0%)	0 (0%)	7 (0.2%)	0 (0%)	2 (0.0%)	0 (0%)	11 (0.1%)	1 (0.0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Education, n (%)														
Less than primary school	930 (9.9%)	1351 (22.6%)	2200 (49.3%)	118 (1.2%)	5079 (51.2%)	3207 (62.3%)	3747 (39.5%)	680 (21.2%)	438 (9.1%)	1354 (26.9%)	786 (12.6%)	194 (16.7%)	3131 (36.4%)	170 (3.3%)
Primary school	1400 (14.8%)	1348 (22.5%)	493 (11.0%)	499 (5.2%)	1867 (18.8%)	481 (9.4%)	2051 (21.6%)	789 (24.7%)	1957 (40.7%)	1456 (28.9%)	1351 (21.7%)	404 (34.8%)	3855 (44.8%)	1312 (25.7%)
Secondary school	3084 (32.7%)	1216 (20.3%)	355 (7.9%)	4131 (42.9%)	1809 (18.2%)	397 (7.7%)	1246 (13.1%)	690 (21.6%)	1060 (22.0%)	1824 (36.2%)	2091 (33.6%)	441 (38.0%)	831 (9.7%)	550 (10.8%)
Higher education	2358 (25.0%)	1471 (24.6%)	940 (21.0%)	1961 (20.3%)	1016 (10.2%)	1059 (20.6%)	1777 (18.7%)	609 (19.0%)	1270 (26.4%)	302 (6.0%)	407 (6.5%)	122 (10.5%)	398 (4.6%)	2055 (40.2%)
College/University	1643 (17.4%)	605 (10.1%)	473 (10.6%)	2930 (30.4%)	145 (1.5%)	0 (0%)	658 (6.9%)	397 (12.4%)	54 (1.1%)	104 (2.1%)	773 (12.4%)	0 (0%)	383 (4.5%)	997 (19.5%)
Missing	13 (0.1%)	0 (0%)	6 (0.1%)	0 (0%)	11 (0.1%)	0 (0%)	18 (0.2%)	35 (1.1%)	31 (0.6%)	0 (0%)	808 (13.0%)	0 (0%)	3 (0.0%)	29 (0.6%)
Marital status, n (%)														
Never married	1571 (16.7%)	1362 (22.7%)	661 (14.8%)	1645 (17.1%)	1321 (13.3%)	396 (7.7%)	1838 (19.4%)	860 (26.9%)	894 (18.6%)	1264 (25.1%)	1516 (24.4%)	266 (22.9%)	1778 (20.7%)	2107 (41.2%)
Currently married	7439 (78.9%)	1948 (32.5%)	3293 (73.7%)	6299 (65.3%)	7431 (74.9%)	4721 (83%)	7005 (73.8%)	2021 (63.2%)	2967 (61.7%)	2883 (57.2%)	4286 (69.0%)	606 (52.2%)	6264 (72.8%)	2749 (53.8%)
Separated	29 (0.3%)	653 (10.9%)	21 (0.5%)	185 (1.9%)	43 (0.4%)	45 (0.9%)	23 (0.2%)	12 (0.4%)	171 (3.6%)	436 (8.7%)	34 (0.5%)	92 (7.9%)	34 (0.4%)	13 (0.3%)
Divorced	75 (0.8%)	19 (0.3%)	58 (1.3%)	153 (1.6%)	165 (1.7%)	9 (0.2%)	60 (0.6%)	46 (1.4%)	80 (1.7%)	37 (0.7%)	128 (2.1%)	62 (5.3%)	55 (0.6%)	64 (1.3%)
Widowed	284 (3.0%)	453 (7.6%)	432 (9.7%)	1347 (14.0%)	780 (7.9%)	423 (8.2%)	566 (6.0%)	251 (7.8%)	379 (7.9%)	413 (8.2%)	249 (4.0%)	120 (10.3%)	469 (5.5%)	136 (2.7%)
Cohabiting	25 (0.3%)	1554 (25.9%)	0 (0%)	6 (0.1%)	164 (1.7%)	0 (0%)	0 (0%)	0 (0%)	280 (5.8%)	1 (0.0%)	0 (0%)	11 (0.9%)	1 (0.0%)	16 (0.3%)
Missing	5 (0.1%)	2 (0.0%)	2 (0.0%)	4 (0.0%)	23 (0.2%)	0 (0%)	5 (0.1%)	10 (0.3%)	39 (0.8%)	6 (0.1%)	3 (0.0%)	4 (0.3%)	0 (0%)	28 (0.5%)
WHODAS subdomains, mean (SD)														
Understanding and communication	1.1 (1.6)	1.2 (2)	1.4 (2.3)	1.3 (2.3)	1.3 (2.2)	1.3 (2)	2.4 (2.7)	0.9 (1.8)	1 (1.7)	0.4 (1)	0.6 (1.3)	1.2 (2)	1.2 (2.1)	1.4 (2.1)
Missing	6 (0.06%)	12 (0.2%)	6 (0.13%)	66 (0.68%)	80 (0.81%)	18 (0.35%)	151 (1.59%)	146 (4.56%)	13 (0.27%)	11 (0.22%)	0 (0%)	6 (0.52%)	5 (0.06%)	29 (0.57%)
Getting around	0.2 (0.8)	0.5 (1.2)	0.9 (1.6)	1 (1.7)	0.4 (1.1)	1 (1.6)	0.9 (1.6)	0.6 (1.3)	0.5 (1.2)	0.3 (0.8)	0.1 (0.6)	0.8 (1.5)	0.7 (1.4)	0.5 (1.1)
Missing	3 (0.03%)	10 (0.17%)	1 (0.02%)	5 (0.05%)	10 (0.1%)	11 (0.21%)	41 (0.43%)	21 (0.66%)	8 (0.17%)	7 (0.14%)	0 (0%)	5 (0.43%)	4 (0.05%)	17 (0.33%)
Self-care	0.3 (1.1)	0.3 (1.1)	1 (1.8)	1.1 (2.1)	0.4 (1.1)	0.5 (1.5)	0.9 (1.8)	0.8 (1.7)	0.2 (1.1)	0.1 (0.6)	0.1 (0.7)	0.6 (1.7)	1.3 (1.8)	0.4 (1.2)
Missing	10 (0.11%)	153 (2.55%)	10 (0.22%)	640 (6.64%)	267 (2.69%)	6 (0.12%)	2885 (30.38%)	157 (4.91%)	21 (0.44%)	5 (0.1%)	0 (0%)	28 (2.41%)	14 (0.16%)	38 (0.74%)
Getting along with people	0.3 (0.8)	0.4 (1.1)	0.7 (1.5)	0.5 (1.2)	0.6 (1.4)	0.2 (0.8)	0.7 (1.5)	0.5 (1.2)	0.4 (1)	0.2 (0.9)	0.2 (0.7)	0.4 (1)	0.6 (1.4)	0.7 (1.4)
Missing	3236 (34.32%)	1029 (17.18%)	1141 (25.54%)	3667 (38.04%)	2334 (23.51%)	1715 (33.34%)	2522 (26.56%)	1450 (45.31%)	1154 (23.99%)	95 (1.88%)	2114 (34.01%)	483 (41.6%)	2711 (31.52%)	114 (2.23%)
Life activities	0.8 (1.9)	0.6 (1.6)	1.6 (2.6)	1.2 (2.3)	1.3 (2.2)	1.3 (2.7)	1.4 (2.5)	1.3 (2.7)	0.6 (1.6)	0.3 (1.3)	0.2 (1)	1.3 (2.1)	1.1 (2.4)	1.4 (2.1)
Missing	1187 (12.59%)	3068 (51.21%)	2250 (50.37%)	6407 (66.47%)	2333 (23.5%)	1639 (31.86%)	4975 (52.38%)	1063 (33.22%)	2171 (45.14%)	39 (0.77%)	1017 (16.36%)	501 (43.15%)	3886 (45.18%)	295 (5.77%)
Participation in society	0.7 (1.4)	0.5 (1.2)	0.7 (1.3)	1.1 (2)	0.6 (1)	1.5 (1.9)	0.8 (1.4)	0.9 (1.6)	0.7 (1.5)	0.2 (0.8)	0.3 (0.9)	0.6 (1.1)	1 (1.7)	0.5 (1.1)
Missing	120 (1.27%)	18 (0.3%)	714 (15.98%)	2521 (26.15%)	198 (1.99%)	31 (0.6%)	165 (1.74%)	75 (2.34%)	500 (10.4%)	89 (1.77%)	0 (0%)	398 (34.28%)	10 (0.12%)	396 (7.74%)

3.2 | Disability weights

Mean disability weight was 0.12 (SD: 0.05), ranging from 0.17 (SD:0.10) in the Iran dataset to 0.10 (SD: 0.04) in the Singapore dataset. For all countries, data were right-skewed with medians ranging from 0.13 (Iran) to 0.09 (Mexico, Nigeria, and Singapore). Distributions of disability weight per country are shown in Figure 2.

3.3 | Fitting of statistical learning models

Model performance for each of the 10 predictor sets (six for WHODAS 2.0–36 and four for WHODAS 2.0–12) for linear regression and LASSO regression is shown in Table 2. The training and test set had similar model performance, attesting to the robustness of the models and indicating that the models did not overfit the training data. Furthermore, the performance of the linear regression and the LASSO regression was similar (RMSE: 0.040–0.046 and R (Donahue et al., 2018): 0.700–0.754 on the test set). Although the linear regression models demonstrated a similar performance compared to the LASSO models, the resulting models showed some counterintuitive coefficients (i.e. small coefficients with the wrong sign for some items). Furthermore, as LASSO reduced the number of predictors used for mapping (e.g. removing less relevant demographics and most of the country-specific interaction terms), with similar model performance, LASSO regression estimates are presented in Table 3.

Table 3 shows the best performing country-independent (generic) mapping functions for WHODAS 2.0–36 and WHODAS 2.0–12, using individual WHODAS 2.0 items as well as demographic variables. Marital status (for WHODAS 2.0–36 only) and education did not contribute to disability weight, while increasing age and being female were associated with higher disability weight. Looking at the impact of individual items, items D1.1/S6 “Concentrating on something for 10 minutes” was shown to have the largest impact on disability weight, followed by D3.1/S8 “Analysing and finding solutions to problems in day-to-day life” and D2.3 “Moving around inside your house” (WHODAS 2.0–36 only).

3.4 | Sensitivity analyses

Limiting the analysis to records with no missing data resulted in a dataset with a total of 21,302 respondents (23.2% of the sample used in the main analysis). Model performance for this subsample was lower than in the main analyses, with an R^2 of 0.645 using WHODAS 2.0–36 and 0.607 using WHODAS 2.0–12. The RMSE for the models in this subsample cannot easily be compared to the RMSE for the models in the main analysis, as the dependent variable in the completers only set had a smaller range. Imputing missing WHODAS domain scores using the mean of those domain scores for other respondents instead of the mean of other domain scores within the same respondent resulted in similar results compared to the main

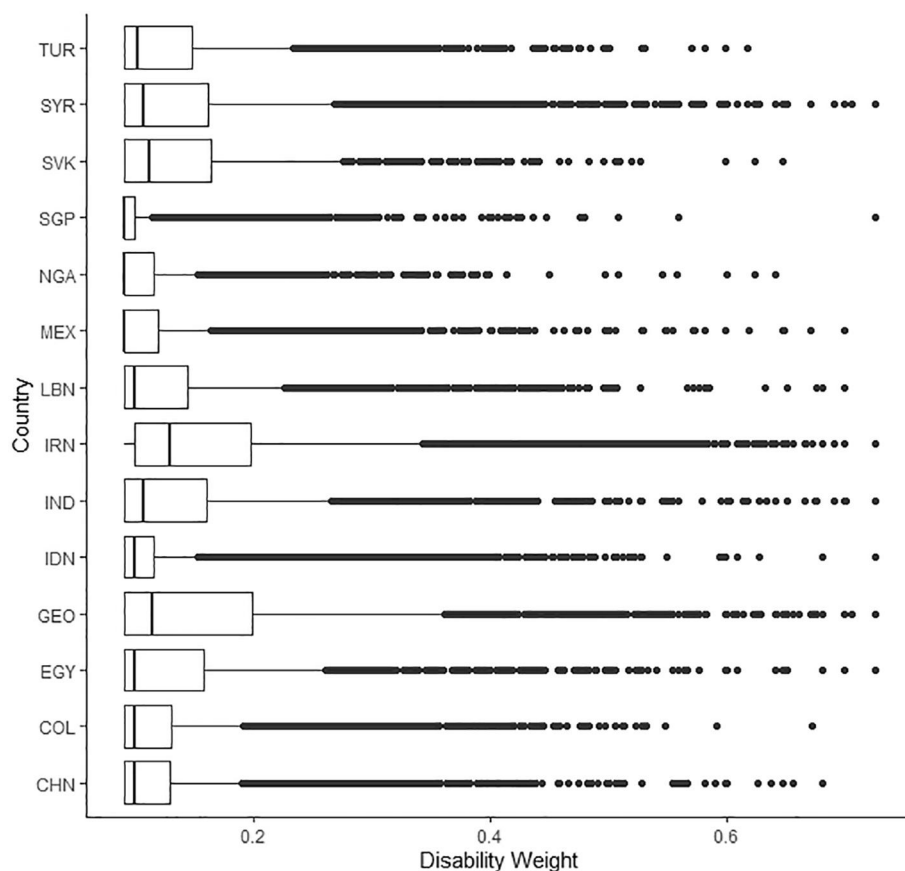


FIGURE 2 Boxplot presenting the distribution of disability weights per country

TABLE 2 Model performance of the various statistical learning models predicting disability weights

Model nr.	WHODAS Version	Method	Predictors included in model/algorithm	RMSE	R (Donahue et al., 2018)	RMSE (Test set) ^b	R (Donahue et al., 2018) (Test set) ^c
1	36-Item	Linear regression	Individual items	0.04	0.743		
2	36-Item	Linear regression	Individual items & demographics^a	0.04	0.743	0.043	0.739
3	36-Item	Linear regression	All six domain scores	0.045	0.676		
4	36-Item	Linear regression	All six domain scores & demographics ^a	0.045	0.676		
5	36-Item	Linear regression	Individual items, demographics ^a & country dummy	0.04	0.743		
6	36-Item	Linear regression	Individual items, demographics^a, country dummy, all country interactions	0.043	0.721	0.042	0.754
7	12-Item	Linear regression	Individual items	0.044	0.701		
8	12-Item	Linear regression	Individual items & demographics^a	0.043	0.705	0.047	0.700
9	12-Item	Linear regression	Individual items, demographics ^a & country dummy	0.043	0.705		
10	12-Item	Linear regression	Individual items, demographics^a, country dummy, all country interactions	0.046	0.682	0.045	0.721
11	36-Item	LASSO regression	Individual items	0.04	0.743		
12	36-Item	LASSO regression	Individual items & demographics^a	0.04	0.743	0.044	0.738
13	36-Item	LASSO regression	All six domain scores	0.045	0.676		
14	36-Item	LASSO regression	All six domain scores & demographics ^a	0.045	0.676		
15	36-Item	LASSO regression	Individual items, demographics ^a & country dummy	0.04	0.743		
16	36-Item	LASSO regression	Individual items, demographics^a, country dummy, all country interactions	0.041	0.742	0.043	0.747
17	12-Item	LASSO regression	Individual items	0.044	0.701		
18	12-Item	LASSO regression	Individual items & demographics^a	0.043	0.704	0.047	0.700
19	12-Item	LASSO regression	Individual items, demographics ^a & country dummy	0.043	0.705		
20	12-Item	LASSO regression	Individual items, demographics^a, country dummy, all country interactions	0.044	0.704	0.046	0.712

Note: Per WHODAS version, per statistical learning method, and for the models with and without country information, the best performing model is bold faced.

^aDemographic variables include: age, gender, educational level, and marital status.

^bRoot-mean-squared error for each model predicting disability weights using WHODAS responses on the test set.

^cR-squared for each model predicting disability weights using WHODAS responses on the test set.

analyses with the alternative analyses leading to slightly lower performance metrics with an R^2 of 0.738 compared to 0.743 (WHODAS 2.0–36) and 0.692 compared to 0.705 (WHODAS 2.0–12) and comparable RMSEs compared to the base case analyses. Likewise, imputing missing WHODAS domain scores using kNN imputation resulted in similar results compared to the main analyses with an R^2 of 0.742 compared to 0.743 (WHODAS 2.0–36) and 0.701 compared to 0.705 (WHODAS 2.0–12) and comparable RMSE values. All

coefficients had similar signs as in the main analyses. For the country-specific models, see Table 4, performance varied from an R^2 of 0.593 (WHODAS 2.0–36; Nigeria) and 0.523 (WHODAS 2.0–12; Nigeria) to 0.811 (WHODAS 2.0–36; Georgia) 0.794 (WHODAS 2.0–12; Georgia). When comparing model performance of the generic models (i.e. for WHODAS 2.0–36 and WHODAS 2.012) to the country-specific models on the country-specific test sets, the generic model performed equally well with only minor differences in model

TABLE 3 Mapping function for WHODAS 2.0–36 and WHODAS 2.0–12 with demographics based on LASSO regression

Predictor	Model 12 (WHODAS 2.0–36) ^{a,b}	Model 18 (WHODAS 2.0–12) ^{a,b}
Intercept	0.1344	0.1344
<i>Items in original WHODAS 2.0–36/WHODAS 2.0–12</i>		
D1.1/S6	0.0273	0.0295
D1.4/S3	0.0011	0.0028
D1.5	-	NA
D1.6	-	NA
D2.2	0.0087	NA
D2.3	0.0121	NA
D3.1/S8	0.0119	0.0161
D3.2/S9	0.0042	0.0081
D3.4	0.0006	NA
D4.2/S11	0.0020	0.0016
D4.3	0.0001	NA
D4.5	0.0006	NA
D5.1/S2	0.0042	0.0119
D5.3	0.0072	NA
D5.5/S12	0.0029	0.0077
D5.7	0.0000	NA
D6.1/S4	0.0067	0.0097
D6.6	0.0010	NA
D6.7	0.0039	NA
<i>Demographic variables</i>		
Age	0.0016	0.0047
Gender (male)	-0.0003	-0.0014
Education	-	-
Marital status (widowed)	-	0.003

Note: Variables excluded from the model by the LASSO procedure are indicated with a '-'

^aFor model specifications see Table 2.

^bAll WHODAS items are converted to a 0–4 scale.

performance between the generic- and country-specific models (see Table 4), favouring the generic model for some countries and the country-specific models for other countries, suggesting that the generic model can be robustly used to estimate DW for all countries. Lastly, training models on data from 13 countries and assessing it on the 14th country resulted in similar model performance to that of the generic model (see Table 4). The country-specific models will be made available upon request to the corresponding author.

3.5 | Example

The mapping functions presented in Table 3 are straightforward to use for converting WHODAS 2.0 scores into disability weight estimates. As a hypothetical example, assume there is a trial in which patients

treated for moderate depression are randomized to either care as usual or care as usual plus additional treatment. In both arms, patients are aged 40 on average at baseline and 50% of patients is female. WHODAS 2.0–36 is administered and patients in both arms score a two on every item of the WHODAS 2.0 (scaled to 0–4 in concordance with the WHODAS scoring manual). This means that at baseline, average disability weight in both groups is 0.38725 (see equation 1). Assume the items D1.1 (*Concentrating on doing something for ten minutes*), D4.2 (*Maintaining a friendship*), D5.1 (*Taking care of your household responsibilities*), D5.3 (*Getting all the household work done that you needed to do*), D5.5 (*Your day-to-day work/school*), D5.7 (*Getting all the work done that you need to do*) and D6.1 (*How much of a problem did you have in joining in community activities (for example, festivities, religious or other activities) in the same way as anyone else can*) are positively impacted by the treatment, improving from 2 to 1.5 in the care as usual

TABLE 4 Comparison of performance for the generic and country-specific models and the model using only data from other countries

Model 12 (WHODAS 2.0–36) ^a	Generic model – R-squared	Country-specific model – R-squared	'Other-countries' model – R-squared ^b
China (N = 9,486)	0.670	0.697	0.721
Colombia (N = 8,158)	0.687	0.690	0.677
Egypt (N = 4,490)	0.793	0.799	0.784
Georgia (N = 9,847)	0.804	0.791	0.818
Indonesia (N = 9,994)	0.627	0.651	0.669
India (N = 5,144)	0.770	0.768	0.771
Iran (N = 9,718)	0.737	0.722	0.742
Lebanon (N = 3,246)	0.803	0.807	0.795
Mexico (N = 4,813)	0.634	0.669	0.707
Nigeria (N = 5,108)	0.596	0.572	0.679
Singapore (N = 6,216)	0.748	0.763	0.835
Slovakia (N = 1,183)	0.805	0.794	0.805
Syria (N = 9,344)	0.741	0.747	0.738
Turkey (N = 5,207)	0.594	0.544	0.599
Model 18 (WHODAS 2.0–12)			
China (N = 9,486)	0.605	0.666	0.667
Colombia (N = 8,158)	0.623	0.636	0.634
Egypt (N = 4,490)	0.766	0.772	0.765
Georgia (N = 9,847)	0.785	0.775	0.786
Indonesia (N = 9,994)	0.585	0.620	0.640
India (N = 5,144)	0.697	0.702	0.701
Iran (N = 9,718)	0.702	0.682	0.703
Lebanon (N = 3,246)	0.788	0.802	0.775
Mexico (N = 4,813)	0.633	0.638	0.704
Nigeria (N = 5,108)	0.526	0.496	0.635
Singapore (N = 6,216)	0.683	0.748	0.764
Slovakia (N = 1,183)	0.774	0.769	0.768
Syria (N = 9,344)	0.697	0.704	0.692
Turkey (N = 5,207)	0.540	0.586	0.539

^aFor model specifications see Table 2.

^bPerformance of models trained on data from the 13 other countries.

group and from 2 to one in the care as usual plus additional treatment group one year after baseline. This would improve disability weights by an estimated $0.38725 - 0.2942 = 0.09305$ in the additional treatment group and by an estimated $0.38725 - 0.3400 = 0.0456$ in the care as usual group. These improvements in disability weight could then be entered into calculation of QALYs gained.

$$\begin{aligned}
 \text{Disability weight} = & 0.1344 + 0.0273 * 2 + 0.0011 * 2 + 0.0087 * 2 \\
 & + 0.0121 * 2 + 0.0119 * 2 + 0.0042 * 2 + 0.0006 * 2 + 0.0020 * 2 \\
 & + 0.0001 * 2 + 0.0006 * 2 + 0.0042 * 2 + 0.0072 * 2 + 0.0029 * 2 \\
 & + 0.0067 * 2 + 0.0010 * 2 + 0.0039 * 2 + 0.0016 * 40 - 0.0003 * 0.50
 \end{aligned}
 \tag{1}$$

4 | CONCLUSION AND DISCUSSION

This study developed a generic mapping algorithm converting WHODAS 2.0 items into disability weights, using MCSS-data of 14 countries. By exploring various model specifications, using both linear regression and LASSO regression, we found good model performances (with $R^2 > 0.70$ on the test set, including for models not using country-specific information. This shows that it is possible to map WHODAS 2.0 scores to disability weights using a simple and usable country-independent mapping function.

A review by Mukuria et al. (2019) has shown a substantial increase in the number of mapping studies to predict instrument-specific health-state values from a health-related quality of life measure (Mukuria et al., 2019). Ordinary least squares models were shown to be the most common approach for these mappings (used $\geq 75\%$ times within each preference-based measure) (Mukuria et al., 2019). Moreover, Mukuria et al. state that 'the appropriateness of mapping functions relies on assessment of applicability for the context while appropriateness of methods relies on the target outcome measure', indicating that comparison of mapping studies should be done cautiously. Similar to the mapping study presented in this paper, Keetharuth and Rowen, 2020 developed a mapping function to predict Recovering Quality of Life Utility Index (a patient-reported mental health-specific preference-based measure) scores from the Health of Nation Outcomes Scale scores (clinician-reported measure). That study, however, resulted in notably lower (adjusted) R-squared values ranging from 0.180 to 0.491.

4.1 | Strengths

Strengths of this study include the use of data from 14 countries in large nationally representative samples (non-institutionalized adults living in private households), with a total of 88,234 usable observations to train and test the mapping functions. Moreover, data splitting (i.e. to create a train and test dataset) and cross-validation was performed for testing robustness of the models and to prevent overfitting. Lastly, various model specifications were explored using a mix of demographical predictors and WHODAS-items/domain level predictors. In addition, country-specific models were estimated in order to evaluate to what extent a generic, country-independent, mapping function is able to achieve similar performance and thus to what extent the generic mapping function is transferable to different countries.

4.2 | Limitations

Results of this study should be viewed in light of the following limitations. First, the 14 countries which were used are not necessarily representative for the whole world, compromising generalizability of our findings. For example, no Western European countries were included in the study. In addition, we had to remove 4.1% of our data for having a missing outcome, further limiting generalizability. Second, disability weights were estimated indirectly using a health state valuation function based on a set of six core domain levels and were not directly assessed, which could have resulted in a lower performance of the mapping function than could have been obtained otherwise. Third, missing values in WHODAS items were imputed by using the mean of the remaining items in the domain of each respondent, and with the mean of available domain scores for that respondent if no items were available within a domain. This method was deemed most practical, as it does not rely on information from

other patients and is therefore applicable to single patients. A method like *k*-Nearest Neighbours imputation could perhaps be more appropriate, but it would also be more difficult to apply for practitioners using this mapping function. The alternative imputation strategy used in the sensitivity analysis did not alter results. Fourth, as our sample consisted of nationally representative samples (for non-institutionalized adults living in private households), the data contained a relatively small range of disability weights, as respondents were generally in good health. Although this is typical for general population studies, it means that the mapping function is primarily applicable to less severe (patient) groups. Fifth, this study applied a relatively simple machine learning method (i.e. LASSO regression) in order to maximize interpretability and ease of use of the results. Higher performance could be achieved using more flexible statistical learning methods such as gradient boosting algorithms, but at the expense of interpretability (i.e. this would lead to what is often referred to as a "black box") (Watson et al., 2019). Sixth, predictions on an individual level can still exhibit quite some variance (the mean disability was 0.13 with a SD of 0.08). However, the value of the current mapping functions lies primarily in either group estimates of disability weights for specific conditions and/or considering incremental disability weight as a result of improved functioning according to WHODAS 2.0-36 or WHODAS 2.0-12. Seventh, there is a minor difference between the wording of item D1.1. of the WHODAS-2.0 and the corresponding item 2006 in the MCSS questionnaire. In the MCSS, this item was asked as "Overall in the last 30 days how much difficulty did you have with concentrating or remembering things?", instead of "In the past 30 days, how much difficulty did you have in concentrating on doing something for ten minutes?". However, given that the difference is minor, we do not expect this discrepancy to be influential. Eighth, although comparisons of performance of the generic model with the country-specific models indicated that the generic mapping function is transferable to different countries, model performance is not equally good for each individual country. It is difficult to explain why model performance is better in some countries than others. Users of our models should therefore consider both country-specific performance (Table 4) as well as country-specific population characteristics (Table 1) to ensure proper application of our models. Ninth, our choice for a main effects model assuming a normal distribution for estimating health state valuations results in more accurate estimates for individuals with VAS values in the mid-range (25-75 on a 0-100 scale), but it will be less accurate for individuals outside that range. Tenth, as previously published (Üstün et al., 2001), missing rates at the respondent and item levels vary considerably between countries. In their study, Üstün et al., 2001 compared the different administration modes and questionnaire types included in the MCSS-study (i.e. in addition to the more extensive face-to-face household survey that was used for the current study) and concluded that missing rates at the respondent and item levels were higher in the full-length survey than for the brief survey. Moreover, they concluded that the respondent level missing data was highest in the full-length survey, likely due to the length of the full version. However, it is also concluded that reliability of the

responses was not correlated with missingness or with representativeness for the respective countries. Hence, it is hypothesized that the variation among the proportions of missing values both within and between countries does not impact generalizability of our results.

Lastly, our mapping function converts WHODAS 2.0 into disability weights. From a policy-making perspective, a conversion to utilities rather than disability weights is also of particular interest, as it facilitates expressing outcomes in terms of QALYs. As utilities are sometimes derived from disability weights by taking one minus the disability weight, this is still possible with the current mapping function, but it comes with the additional limitation that this results in an estimate of utility and not in an exact utility value.

ACKNOWLEDGEMENTS

Financial or other support: dr Lokkerbol was supported by The Netherlands Organisation for Health Research and Development (ZonMw) Mental Healthcare Fellowship, grant # 636320003.

CONFLICT OF INTEREST

Dr. Chatterji and Dr. Chisholm are staff members of the World Health Organization. The authors alone are responsible for the views expressed in this publication and they do not necessarily represent the decisions, policy or views of the World Health Organization. - In the past 3 years, Dr. Kessler was a consultant for Datastat, Inc, Sage Pharmaceuticals, and Takeda.

DATA AVAILABILITY STATEMENT

Data availability Due to privacy regulations the data cannot be made available. The syntax of the analyses is available upon request at the corresponding author.

ORCID

Ben F. M. Wijnen  <https://orcid.org/0000-0001-7993-1905>

Somnath Chatterji  <https://orcid.org/0000-0002-3085-6550>

Ronald C. Kessler  <https://orcid.org/0000-0003-4831-2305>

REFERENCES

- Afzali, H. H. A., Karnon, J., & Merlin, T. (2013). Improving the accuracy and comparability of model-based economic evaluations of health technologies for reimbursement decisions: A methodological framework for the development of reference models. *Medical Decision Making*, 33(3), 325–332.
- Anderson, C., Laubscher, S., & Burns, R. (1996). Validation of the Short Form 36 (SF-36) health survey questionnaire among stroke patients. *Stroke*, 27(10), 1812–1816.
- Balestroni, G., & Bertolotti, G. (2012). EuroQol-5D (EQ-5D): An instrument for measuring quality of life. *Monaldi Archives for Chest Disease*, 78(3), 155–159.
- Brazier, J., Roberts, J., & Deverill, M. (2002). The estimation of a preference-based measure of health from the SF-36. *Journal of Health Economics*, 21(2), 271–292.
- Buttorff, C., Hock, R. S., Weiss, H. A., Naik, S., Araya, R., Kirkwood, B., Chisholm, D., & Patel, V. (2012). Economic evaluation of a task-shifting intervention for common mental disorders in India. *Bulletin of the World Health Organization*, 90, 813–821.
- Chambers, J. (2008). *Software for data analysis: Programming with R*. Springer Science & Business Media.
- Collins, G. S., Reitsma, J. B., Altman, D. G., & Moons, K. G. M. (2015). Transparent reporting of a multivariable prediction model for individual Prognosis or Diagnosis (TRIPOD) the TRIPOD statement. *Circulation*, 131(2), 211–219.
- Devlin, N. J., & Brooks, R. (2017). EQ-5D and the EuroQol group: Past, present and future. *Applied Health Economics and Health Policy*, 15(2), 127–137.
- Dolan, P. (1997). Modeling valuations for EuroQol health states. *Medical Care*, 35, 1095–1108. Published online.
- Donahue, K., Reid, A., Baxley, E., Carter, C., Carek, P. J., Robinson, M., & Newton, W. P. (2018). Triple aim is triply tough: Can you focus on three things at once? *Family Medicine*, 50(3), 179–187.
- Drummond, M. F., Sculpher, M. J., Claxton, K., Stoddart, G. L., & Torrance, G. W. (2015). *Methods for the economic evaluation of health care programmes*. Oxford university press.
- Field, A. (2013). *Discovering statistics using IBM SPSS statistics*. Sage.
- Garin, O., Ayuso-Mateos, J. L., Almansa, J., Nieto, M., Chatterji, S., Vilagut, G., Alonso, J., Cieza, A., Svetskova, O., Burger, H., Racca, V., Francescutti, C., Vieta, E., Kostanjsek, N., Raggi, A., Leonardi, M., & Ferrer, M. (2010). Validation of the “World Health Organization disability assessment schedule, WHODAS-2” in patients with chronic diseases. *Health and Quality of Life Outcomes*, 8(1), 51.
- Homedes, N. (1996). *The disability-adjusted life year (DALY) definition, measurement and potential use*. The World Bank.
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning* (Vol. 112). Springer.
- James, S. L., Abate, D., Abate, K. H., Abay, S. M., Abbafati, C., Abbasi, N., Abbastabar, H., Abd-Allah, F., Abdela, J., Abdelalim, A., Abdollahpour, I., Abdulkader, R. S., Abebe, Z., Abera, S. F., Abil, O. Z., Abraha, H. N., Abu-Raddad, L. J., Abu-Rmeileh, N. M. E., Accrombessi, M. M. K., & Murray, C. J. L. (2018). Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990–2017: A systematic analysis for the global burden of disease study 2017. *Lancet*, 392(10159), 1789–1858.
- Keetharuth, A., & Rowen, D. (2020). *Mapping the health of nation outcomes scale (HoNOS) onto the recovering quality of life utility Index*. ReQoL-UI) Published online 2020.
- Kuhn, M. (2012). The caret package. In: *R Found Stat Comput Vienna*, URL <https://cranr-project.org/package=caret>. Published online 2012.
- Mukuria, C., Rowen, D., Harnan, S., Rawdin, A., Wong, R., Ara, R., & Brazier, J. (2019). An updated systematic review of studies mapping (or cross-walking) measures of health-related quality of life to generic preference-based measures to generate utility values. *Applied Health Economics and Health Policy*. 17(3):295–313. Published online, 1–19.
- Murray, C. J. L., & Evans, D. B. (2003a). *Health systems performance assessment: Debates, methods and empiricism*. World Health Organization.
- Murray, C. J. L., & Evans, D. B. (2003b). Health systems performance assessment: Goals, framework and overview. In *Health systems performance assessment: Debates, methods and empiricism* (pp. 3–23). World Health Organization.
- Salomon, J. A., Haagsma, J. A., Davis, A., de Noordhout, C. M., Polinder, S., Havelaar, A. H., Cassini, A., Devleeschauwer, B., Kretzschmar, M., Speybroeck, N., Murray, C. J. L., & Vos, T. (2015). Disability weights for the global burden of disease 2013 study. *Lancet Glob Heal*, 3(11), e712–e723.
- Sayak, P. (2018). *Towards preventing overfitting: Regularization*. <https://www.datacamp.com/community/tutorials/towards-preventing-overfitting-regularization>. Published Accessed 8 May 2019.
- Schmidt, H., Gostin, L. O., & Emanuel, E. J. (2015). Public health, universal health coverage, and sustainable development goals: Can they coexist? *Lancet*, 386(9996), 928–930.

- Üstün, T. B., Chatterji, S., Kostanjsek, N., Rehm, J., Kennedy, C., Epping-Jordan, J., Saxena, S., von Korf, M., & Pull, C. (2010). Developing the World Health Organization disability assessment schedule 2.0. *Bulletin of the World Health Organization*, 88, 815–823.
- Üstün, T. B., Chatterji, S., Villanueva, M. (2001). *WHO multi-country survey study on health and responsiveness*. World Heal Organ. Published online.
- Üstün, T. B., Kostanjsek, N., Chatterji, S., & Rehm, J. (2010). *Measuring health and disability: Manual for WHO disability assessment Schedule WHODAS 2.0*. World Health Organization.
- vanVelden, M. E., Severens, J. L., & Novak, A. (2005). Economic evaluations of healthcare programmes and decision making. *Pharmacoeconomics*, 23(11), 1075–1082.
- Walker, D. G., Hutubessy, R., & Beutels, P. (2010). WHO Guide for standardisation of economic evaluations of immunization programmes. *Vaccine*, 28(11), 2356–2359.
- Watson, D. S., Krutzinna, J., Bruce, I. N., Griffiths, C. E., McInnes, I. B., Barnes, M. R., & Floridi, L. (2019). Clinical applications of machine learning algorithms: Beyond the black box. *BMJ*, 364, 1886.
- Wiens, J., & Shenoy, E. S. (2018). Machine learning for healthcare: On the verge of a major shift in healthcare epidemiology. *Clinical Infectious Diseases*, 66(1), 149–153.
- Wijnen, B. F. M., Mosweu, I., Majoie, M. H. J. M., Ridsdale, L., de Kinderen, R. J. A., Evers, S. M. A. A., & McCrone, P. (2018). A comparison of the responsiveness of EQ-5D-5L and the QOLIE-31P and mapping of QOLIE-31P to EQ-5D-5L in epilepsy. *The European Journal of Health Economics*, 19(6), 861–870.
- Williams, I., McIver, S., Moore, D., & Bryan, S. (2008). The use of economic evaluations in NHS decision-making: A review and empirical investigation. *Health Technology Assessment*, 12(7), 1–175.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Lokkerbol, J., Wijnen, B. F. M., Chatterji, S., Kessler, R. C., & Chisholm, D. (2021). Mapping of the World Health Organization's Disability Assessment Schedule 2.0 to disability weights using the Multi-Country Survey Study on Health and Responsiveness. *International Journal of Methods in Psychiatric Research*, 30(3), e1886. <https://doi.org/10.1002/mpr.1886>