

Q-squared as a mapping model performance metric: comparison of models mapping to EQ-5D in gastrointestinal disease

Kristian Mallon

mallonkristian21@gmail.com

Atlantic Technological University

Rositsa Koleva-Kolarova

University of Oxford

Richeal Burns

Atlantic Technological University

Yaling Yang

University of Oxford

Helen Dakin

University of Oxford

Research Article

Keywords: Mapping, Q-squared, predictive accuracy, EuroQoL EQ-5D, Gastrointestinal Disease; quality of life

Posted Date: September 29th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10617892/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Q-squared as a mapping model performance metric: comparison of models mapping to EQ-5D in gastrointestinal disease

Authors:

Kristian Mallon, PhD^{1,2,3}, Rositsa Koleva-Kolarova, PhD⁴, Richeal Burns, PhD^{1,2}, Yaling Yang, PhD⁵, Helen Dakin, DPhil⁴

¹ Department of Health and Nutritional Science, Atlantic Technological University, Sligo, Ireland.

² The Health and Biomedical Research Centre (HEAL), Atlantic Technological University, Sligo, Ireland.

³ Higher Colleges of Technology, Abu Dhabi, United Arab Emirates.

⁴ Nuffield Department of Population Health, University of Oxford, Oxford, UK.

⁵ Nuffield Department of Primary Care Health Sciences, University of Oxford, Oxford, UK.

Corresponding author: Dr Kristian P. Mallon, Department of Health and Nutritional Science, Atlantic Technological University, Sligo, Ireland; The Health and Biomedical Research Centre (HEAL), Atlantic Technological University, Sligo, Ireland.

Ash Lane, Sligo, F91 YW50, Ireland.

Tel: + 353 (0)71 91 55222.

Email: mallonkristian21@gmail.com

Abstract

Background:

Model performance in mapping studies between quality of life measures is often assessed using metrics such as root mean square error (RMSE), mean squared error (MSE), and R-squared. The Q-squared (Q^2) statistic has recently been used as a measure of prognostic accuracy, but has not previously been applied to assess model performance in quality of life mapping studies. We aimed to estimate Q-squared for published mapping algorithms in gastrointestinal (GI) disease and evaluate its usefulness as a metric of prediction accuracy and model performance.

Methods:

We used the Oxford Health Economics Research Centre (HERC) Database of Mapping Studies to identify published studies estimating or validating algorithms mapping quality of life measures to the EQ-5D and other target instruments in GI disease populations. Q-squared was calculated from reported statistics as: $1 - \text{RMSE}^2 / \text{SD}^2$. Values of: ≥ 0.75 , 0.50-0.74, 0.25-0.49, 0-0.24, and < 0 were interpreted as substantial, moderate, weak, poor, and negative explanatory power, respectively.

Results:

Fourteen studies were included, encompassing 14,609 individuals with GI disease and 118 mapping functions. Q-squared ranged from -0.69 to 0.91 (median: 0.52). Overall, 50% (59/118) of algorithms showed moderate (0.50-0.74) predictive performance, and 8% (10/118) showed substantial (≥ 0.75) predictive performance. Q-squared in internal validation samples consistently exceeded that in external validation samples: 81% (46/57) of internal validation functions showed moderate/substantial performance, versus 62% (16/26) of external validation functions showing weak performance.

Conclusions:

Q-squared is an informative measure of predictive performance for mapping studies between quality-of-life measures. It can help select between algorithms, benchmark predictive power, or support model averaging, classifying predictive performance into interpretable groups that add value beyond existing metrics.

Keywords: Mapping, Q-squared, predictive accuracy, EuroQoL EQ-5D, Gastrointestinal Disease; quality of life.

Introduction

EQ-5D is the most widely used generic preference-based health-related quality of life (HRQoL) instrument⁽¹⁾ and is recommended for use in economic evaluation reference case by the National Institute for Health and Care Excellence (NICE).⁽²⁾ EQ-5D contains five dimensions which cover generic aspects of HRQoL including mobility, self-care, usual activities, pain and discomfort, and anxiety and depression, and each dimension has 3 (for EQ-5D-3L) and 5 (for EQ-5D-5L) responses. The combination of the 5 dimensions and the responses defines EQ-5D health state which can be used to calculate utility values using the off-the-shelf population value set. The EQ-5D estimates health state preference values or ‘utilities’ that enable the estimation of quality-adjusted life-years (QALYs) within economic evaluations of health technologies.⁽³⁾ Where data on EQ-5D are unavailable, mapping techniques have become increasingly used.⁽⁴⁾ Mapping has been defined as the development and use of a regression model or algorithm to predict values or responses on one HRQoL instrument using data from other measures.⁽⁴⁾ To date, approximately 275 studies have reported 393 algorithms that map a broad range of clinical and disease or condition specific HRQoL instruments across multiple disease areas to EQ-5D.^(5,6)

Model performance in mapping studies is often assessed using metrics such as root mean square error (RMSE), mean squared error (MSE), mean absolute error (MAE), and occasionally R-squared (R^2).⁽⁷⁾ MAE and MSE are useful for selecting the best regression model and comparing prediction accuracy across models estimated on the same dataset. Both measure prediction accuracy and penalise biased models that systematically under- or over-estimate observed values, as well as penalising models with low discrimination that do not correctly identify which patients

have high or low observed values.(8) MSE is often preferred to MAE as it places greater weight on large deviations between predicted and observed utilities.(8) However, MAE and MSE cannot be directly compared between samples as they tend to be larger in samples with a larger standard deviation (SD). R^2 is generally not recommended for comparing mapping models(9) as it captures the strength of the correlation rather than prediction accuracy and does not penalise models with biased predictions.(8) Recently, Q-squared, the proportional reduction in error, has been proposed as an alternative metric for assessing prediction accuracy in any model predicting a continuous outcome.(8) Q-squared is based on the ratio of MSE to the SD, so it measures prediction accuracy (including both bias and discrimination) and can be directly compared across models evaluated on different samples. Although it has been proposed as a useful metric for mapping studies,(8) we are not aware of any mapping studies that have used it.

To evaluate the usefulness and appropriateness of Q-squared in this context, we calculated Q-squared for mapping algorithms reported in studies included in Oxford Health Economics Research Centre (HERC) Database of Mapping Studies and focused on the gastrointestinal (GI) disease area.(5) GI disease was selected to ensure a manageable sample for comparison, whilst also providing sufficient detail to evaluate Q-squared across diverse studies, including multiple source and target instruments and model types. This paper aims to:

- (1) Estimate Q-squared for published mapping algorithms estimated on GI disease populations and evaluate its usefulness and appropriateness as a metric for prediction accuracy and model performance.
- (2) Assess how Q-squared varies by mapping model type, sample size, source-target instrument combination, and validation samples.

Methods

Data Sources

The Oxford Health Economics Research Centre (HERC) Database of Mapping Studies (version 10.0)(5,6,10) was searched to identify published studies mapping HRQoL measures to any version of EQ-5D in GI disease populations.

The HERC Database of Mapping Studies includes key data derived from studies mapping to the EQ-5D, including citation details; the source instrument(s); the clinical area (classified using ICD-10 chapter headings); the number of paired observations of EQ-5D and source instruments used to estimate mapping models; and the types of statistical model/method evaluated.(10) The database is updated regularly as new studies are published.

Identifying relevant studies:

Studies that were already included in the Oxford HERC Database of Mapping Studies version 10(5) were included in our analysis if they:

- Included GI disease populations, defined using International Classification of Diseases-10 (ICD-10)-chapter classifications encompassing diseases of the digestive system, including Irritable Bowel Syndrome, constipation, and Non-alcoholic Steatohepatitis (Chapter XI); and GI cancers including oesophageal, gastric, colon and colorectal (Chapter II). Studies that included patients with a variety of cancers, including GI cancers, were included.
- Published before 17th February 2025
- Used models estimating responses or health state values for EQ-5D-3L or EQ-5D-5L.(6)

- Presented sufficient statistics to allow calculation of Q-squared for the mapping algorithm (estimation, internal validation, or external validation). This required studies to report (a) either MSE or RMSE; and (b) the SD of the target instrument in the same sample.

In addition to identifying primary mapping studies, the HERC Database of Mapping Studies also reports studies conducting external validation of other algorithms, where the external validation study was identified in the main literature searches.⁽⁵⁾ Our study therefore also included external validation studies that were identified in version 10 of the HERC Database of Mapping Studies⁽⁵⁾ and externally validated algorithms meeting our inclusion criteria on GI populations.

Data extraction

We used data from the HERC Database of Mapping Studies, including study year, disease, source and target instruments, and statistical model type (Supplementary Table 1). Data on performance metrics were extracted to manually calculate Q-squared, including SD and either MSE or RMSE and SD. These study characteristics and metrics were extracted for all mapping algorithms and for samples (estimation, internal validation, and external validation). Since most studies reported multiple mapping models, often with several specifications per model, this provided multiple Q-squared values for each study. Subgroup-level data were also extracted where reported. The data extraction table is presented in Supplementary Table 2 (Data extraction table).

The estimation sample was defined as the sample on which the algorithm was estimated: e.g. the sample on which a regression model for mapping was estimated.^(9,11) The internal validation sample was defined as a sub-sample from the same data source with the estimation sample; for

example, by randomly selecting a certain percentage of patients in a study to conduct model validation where the data of the remaining patients are used to estimate the mapping model.(9,11) Internal validation is often used to avoid overfitting when selecting the best model. However, an internal validation sample is likely to be very similar to the estimation sample. The external validation sample was defined as a sample of patients collected independently of the estimation sample, e.g., from a different study.(9) We hypothesised that Q-squared in internal validation samples would be lower than in estimation samples (due to avoiding overfitting), but higher than in external validation samples (which differs more from the estimation sample).

Calculation of Q-squared

Q-squared, the proportional reduction in error, has previously been used as a metric for assessing the predictive accuracy of an econometric model.(8,12) It has also been used as a criterion for detecting outliers or influential observations, or as a test of prognostic accuracy.(13,14) The upper bound on Q-squared is 1, indicating perfectly predicted outcomes. There is no lower bound for Q-squared, which can also be negative, indicating a very poor, biased model.

Q-squared can be calculated as:

$$Q^2 = 1 - \frac{MSE}{SD^2}$$

Where SD is the sample standard deviation, and MSE is calculated for the sample of interest as:

$$MSE = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$$

where Y_i is the i th observed value (e.g. observed EQ-5D utility) and $\hat{Y}_i$ is the prediction.

In structural equation modelling, the Stone-Geisser Q-squared, or cross-validated redundancy measure, is sometimes used to assess predictive accuracy.(15,16) Like the Q-squared measure used in this paper, the Stone-Geisser Q-squared equals 1-(sum of squares of prediction error ÷ sum of squares of observations).(15) However, the Stone-Geisser Q-squared is specifically estimated using blindfolding of partial least squares models, whereas the measure used here can be applied to predicted and observed values from any regression model or predictive algorithm for any sample.

Many papers report RMSE but not MSE. RMSE is often estimated by statistical packages following ordinary least squares regression. Since $RMSE^2 = MSE$ for the estimation sample in ordinary least squares regression, we calculated Q-squared using RMSE as:

$$Q^2 = 1 - \frac{RMSE^2}{SD^2}$$

Analytical methods and interpretation of Q-squared

We describe and plotted Q-squared across and within all included studies, by source-to-target instrument, sample size, and model type. Higher Q-squared values are interpreted as indicating stronger predictive performance. All analyses are exploratory and intended to describe Q-squared trends; no tests of statistical significance were conducted.

We also explored variation in predictive accuracy by categorising Q-squared values into high, moderate, weak or poor and comparing frequencies and proportions by model type and study characteristics. To categorise predictive performance, we adopted and applied threshold values that have been used to categorise the coefficient of determination (R^2) values in the existing

methodological literature.(17) R^2 values of 0.75, 0.50, and 0.25 are commonly interpreted as cut-offs defining models of substantial, moderate, and weak explanatory power.(17) Similar cut-offs have been proposed for the Stone-Geisser Q-squared .(16)

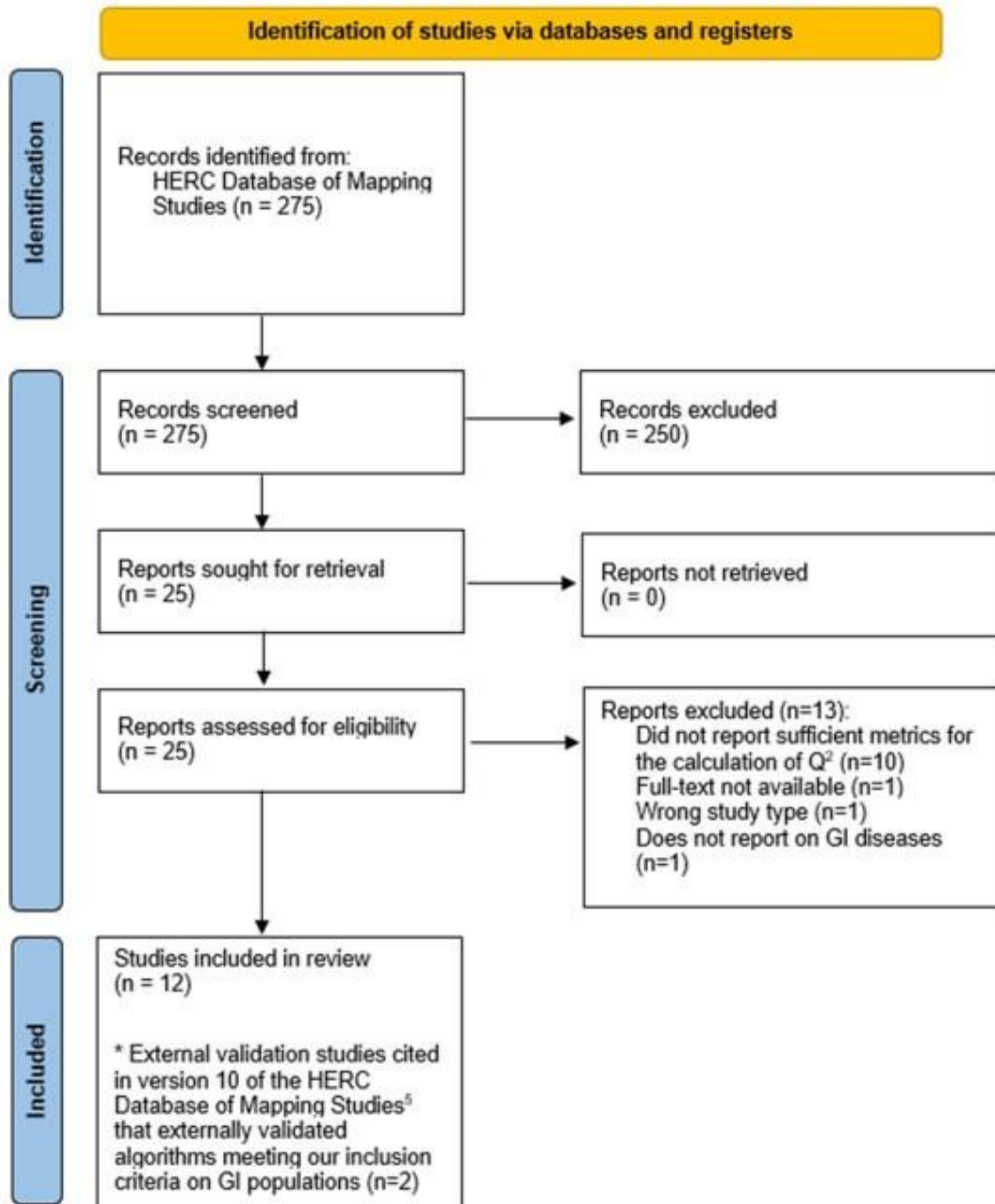
Although these thresholds have not been empirically validated for Q-squared, we applied them as a pragmatic reference point for interpreting a model's predictive performance. We categorised Q-squared values into five bands: substantial (≥ 0.75); moderate (0.50-0.74); weak (0.25-0.49); poor (0-0.24); and negative (< 0). Negative Q-squared values indicate a very poor model where the MSE is larger than the SD; such a model performs worse than a model with just a constant term (which would predict that all patients have the mean value).

Results

Study characteristics

Twelve studies estimating mapping algorithms in patients with GI disease were encompassing 14,609 individuals (Figure 1). Two studies(18,19) externally validating mapping algorithms in GI disease populations were included: the Kontodimopoulos et al. 2009(20) algorithm was validated by Doble and Lorgelly, 2016,(19) while the Marriot et al. 2017(21) algorithm was validated by Woodcock and Doble, 2018.(18) Among the primary studies and the external validation studies included, none reported Q-squared . Ten studies were excluded on account of not reporting statistics enabling manual calculation of Q-squared i.e., SD and either RMSE or MSE.

Figure 1. PRISMA diagram



Data on colorectal cancer were used in nine studies,(18,20–22,22–26) oesophago-gastric cancer(20,21) and constipation,(27,28) in two studies each, and non-alcoholic steatohepatitis,(29) IBS,(30) gastric,(20) gallbladder,(23) liver(23) and pancreatic cancers,(23) each in one study (Table 1).

The source instrument in seven primary studies was the European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire Core 30 (EORTC QLQ-C30).(20–24,26,31) Other source instruments investigated in mapping studies included the Patient Assessment of Constipation Quality of Life (PAC-QOL),(27,28) the Functional Assessment of Cancer Therapy - General (FACT-G),(25) the Irritable Bowel Syndrome Quality of Life (IBS-QOL),(30) and the Chronic Liver Disease Questionnaire for Non-alcoholic Steatohepatitis (CLDQ-NASH).(29)

Estimation sample sizes ranged from 48(20) to 5,421(28) observations (Table 1). Model specifications applied across mapping studies included ordinary least squares (OLS), Tobit, mixture models (adjusted limited dependent variable mixture models [ALDVMM], betamix), response mapping, logit, and censored least absolute deviations (CLAD).

Four of the papers estimating mapping algorithms included internal validation on a sample set aside from the same sample used to estimate the algorithm.(23,27,30,31) One of these papers included external validation in the same paper, using data from a different source from that used to estimate the algorithm.(27) A further two mapping studies had one or more of their mapping algorithms externally validated in a subsequent publication.(20,21)

Characteristics and performance of studies according to Q-squared

The 14 studies (12 primary studies and two external validation studies) provided data allowing the estimation of 118 Q-squared values for 77 mapping algorithms in estimation and validation samples and for different model specifications and subgroups. Q-squared ranged from -0.69 to 0.91. Overall, 8% (10/118) Q-squared values indicated substantial predictive performance (Q-squared: ≥ 0.75), 50% (59/118) moderate (Q-squared: 0.50-0.74), 24% (28/118) weak (0.25-0.49) and 16% (19/118) poor (< 0.25). Two studies reported a negative Q-squared value indicating a biased model: one for an OLS model (-0.69),(27) and one for a CLAD model (-0.01).(30) All data extracted from the included studies, including data used to manually calculate Q-squared, are presented in Supplementary Table 1.

Predictive performance varied between studies and between source/target instruments (Figure 2). Across the 118 Q-squared values from three studies estimating mapping algorithms for EORTC QLQ-C30 to EQ-5D-5L mapping, 97% (30/31) were moderate (Q-squared: 0.50-0.74), and 3% (1/31) were substantial predictive performance (Q-squared: ≥ 0.75). The EORTC QLQ-C30 to EQ-5D-5L model with the highest estimation sample Q-squared (0.73) was the OLS model developed by Perwitasari et al.(24) The EORTC QLQ-C30 to EQ-5D-5L model with the highest internal sample Q-squared (0.76) was the OLS model developed by Yousefi et al.(26)

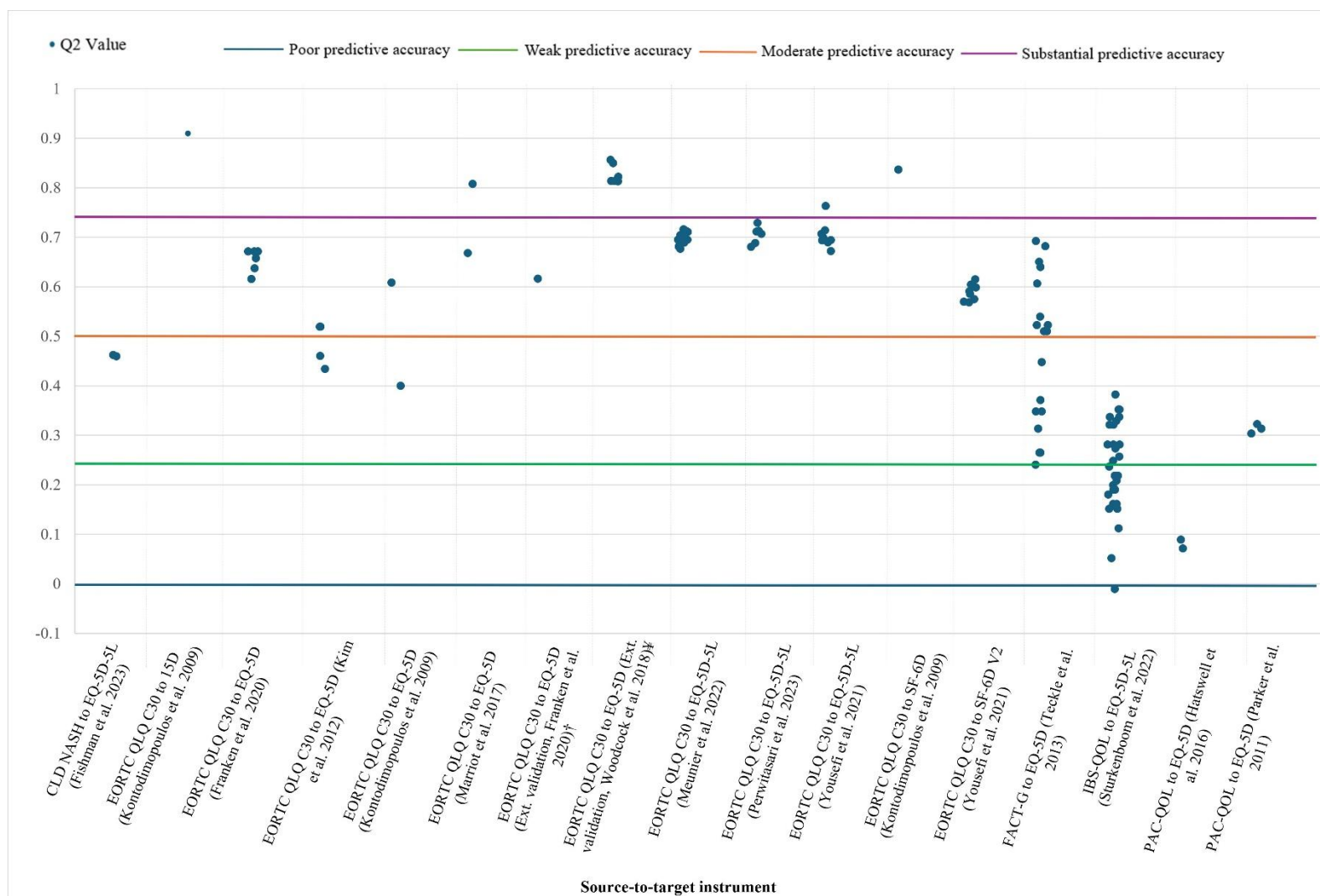
Similar findings were observed in the three studies estimating mapping algorithms and the two external validation studies for EORTC QLQ-C30 to EQ-5D-3L, with 33% (7/21) indicating substantial predictive performance and 52% (11/21) indicating moderate predictive performance. The EORTC QLQ-C30 to EQ-5D-3L model with the highest estimation sample Q-squared (0.81)

was the OLS model developed by Marriot et al.(21) The highest external validation sample Q-squared (0.86) was observed in the study by Woodcock et al.(18)

The poorest predictive performance was observed for the study mapping from the IBS-QOL to EQ-5D-5L, where 53% (16/30) of Q-squared values indicated poor (Q-squared: 0-0.24) predictive performance and the study mapping from PAC-QOL to EQ-5D-3L, where 50% (3/6) of Q-squared were weak (Q-squared: 0.25-0.49) and 50% (3/6) indicated weak predictive performance (Q-squared: 0.25-0.49) and 33% (2/6) were poor (Q-squared : 0-0.24). The lowest Q-squared value was observed when mapping the PAC-QOL to the EQ-5D-3L in a population with opioid-induced constipation (Q-squared = -0.69). Across all mapping functions, an algorithm mapping from the QLQ-C30 to the 15D in a gastric cancer population had the highest predictive accuracy (Q-squared = 0.91), approaching the upper bound of 1.

OLS regression was the most commonly used mapping model approach, with Q-squared values available for 39 algorithm specifications across the included studies, followed by CLAD regression, with 14 algorithm specifications. Beta-regression and separate-equation/three-part models were the least commonly applied, each with only one Q-squared value. Q-squared values varied across model specifications, although no clear pattern emerged (Figure 3).

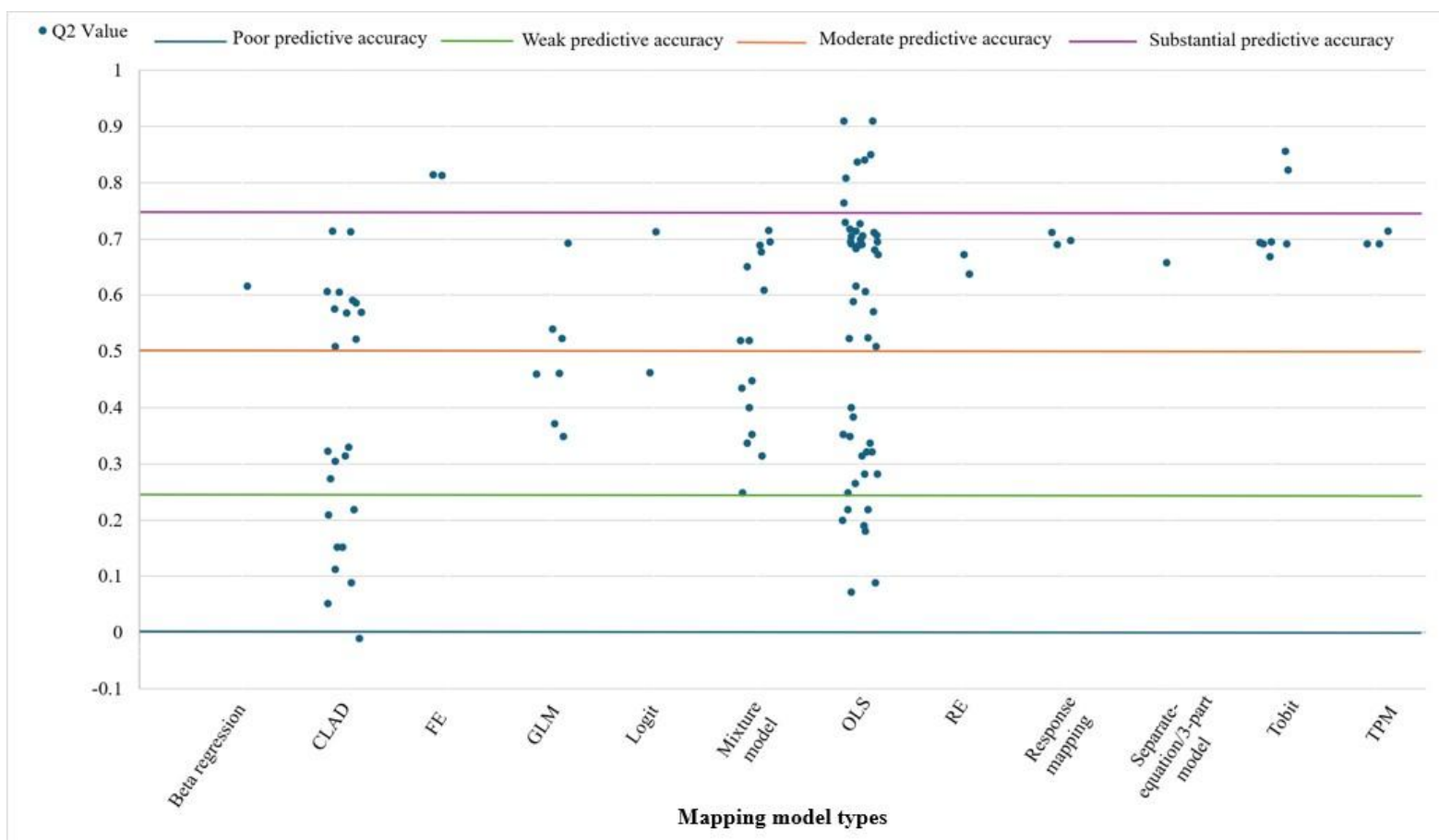
Figure 2. Variation in Q-squared values by study and source-to-target instrument, across sample types (estimation, internal validation and external validation).



Key: †External validation of Marriot et al. (2017) by Franken et al. (2020); ¥External validation of Marriot et al. (2017) by Woodcock et al. (2018).

Abbreviations: CLD-NASH, Chronic Liver Disease Questionnaire for Non-alcoholic Steatohepatitis questionnaire; EORTC QLQ-C30, The European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire Core 30; SF-6D, Short-Form 6 Dimensions; FACT-G, Functional Assessment of Cancer Therapy-General; IBS-QOL, Irritable Bowel Syndrome Quality of Life Instrument (IBS-QOL); PAC-QOL, Patient Assessment of Constipation-Quality

Figure 3. Variation in Q-squared values according to mapping model type, across all sample types (estimation, internal validation and external validation).



Abbreviations: CLAD, Censored Least Absolute Deviations; FE, Fixed Effects; GLM, Generalised Linear Model; OLS, Ordinary Least Squares; RE, Random Effects; TPM, Two-part Model

Mapping functions by study, source-to-target instrument, and validation sample type

Differences in predictive performance were also evident between the estimation sample and internal or external validation samples. Four studies provided sufficient data(23,27,30,31) on two or more validation sample types to enable pairwise comparisons of Q-squared values (Figure 4).

In the study by Hatswell et al.,(27) mapping the PAC-QOL to the EQ-5D, both estimation samples demonstrated poor predictive performance, with Q-squared values of 0.07 and 0.09, respectively. Moreover, external validation of the Parker et al. (2011) algorithm in a similar opioid-induced constipation population yielded a negative Q-squared (Q-squared = -0.69), indicating that the model performed worse than simply predicting the mean utility value.

For Kim et al.(23) mapping the EORTC QLQ-C30 to EQ-5D, both functions demonstrated moderate predictive performance in the estimation sample with Q-squared values of 0.52 for both OLS model specifications. External validation showed weak predictive performance across both functions with Q-squared values of 0.46 and 0.43.

The study by Meunier et al.(31) mapping the EORTC QLQ-C30 to EQ-5D-5L showed the most consistent patterns across all included studies. All six estimation sample functions demonstrated moderate predictive performance with Q-squared values ranging from 0.71 to 0.72. Internal validation Q-squared values were slightly but consistently lower than those for estimation Q-squared, demonstrating moderate predictive performance, with values ranging from 0.68 to 0.70.

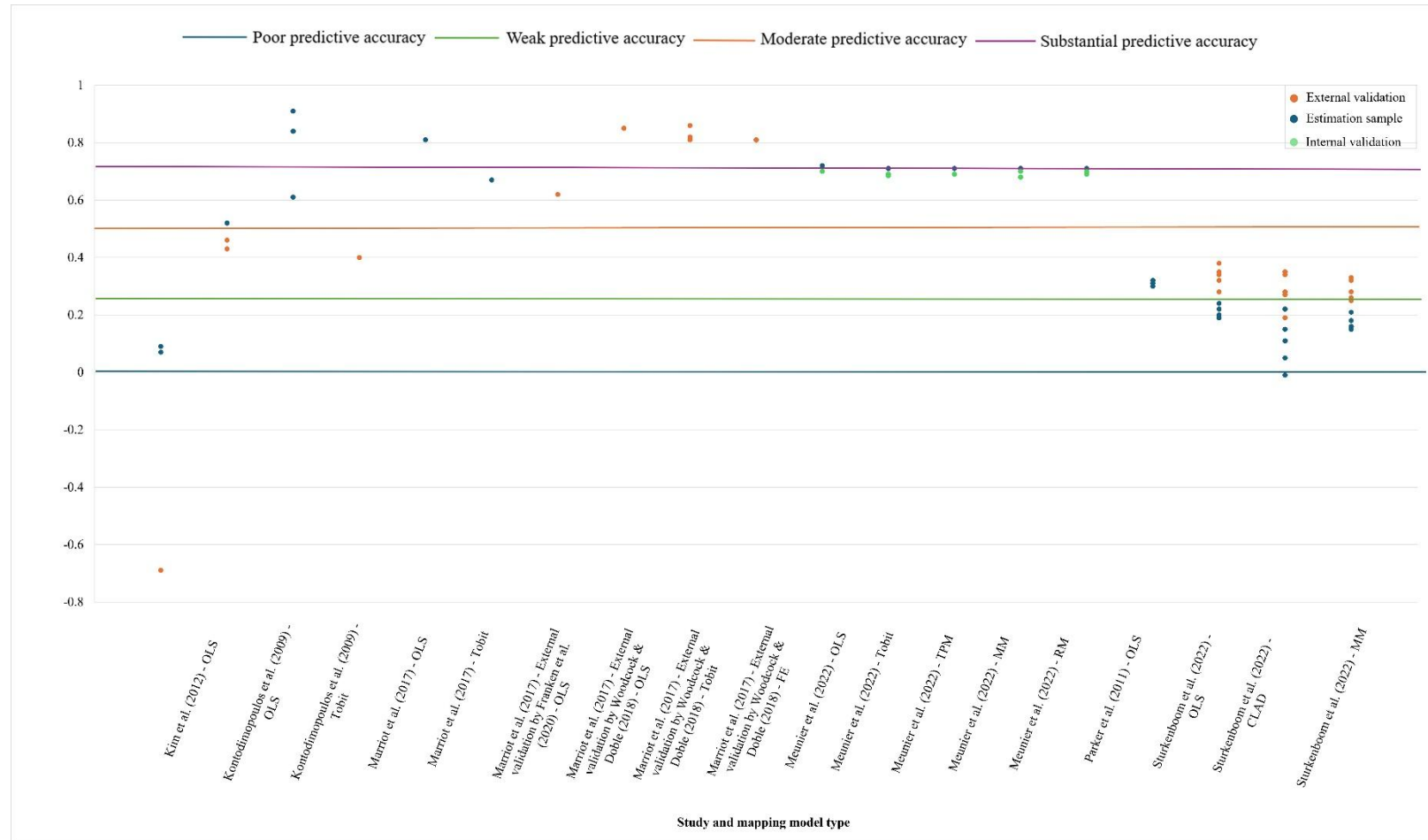
Contrastingly, for Sturkenboom et al.(30) mapping the IBS-QoL to EQ-5D-5L, 93% (14/15) estimation sample Q-squared values were poor (Q-squared: 0-0.24) with one function producing a negative Q-squared value (Q-squared = -0.01).

Sample Size

We examined Q-squared by the sample size of the estimation sample (Supplementary Figure 1). No clear patterns emerged. Among the smallest samples ($n \leq 200$), performance was most variable, with 64% (39/61) demonstrating weak or poor predictive performance. The strongest and most consistent predictive performance was observed in the 501-800 sample size group, which included 35 Q-squared values from two studies.(26,31) All mapping functions demonstrated at least moderate predictive performance, with 33 (94%) within the moderate range in this group.

Additional data on the predictive performance of mapping algorithms in the estimation sample by study characteristics are presented in Supplementary Table 3 (Predictive performance of mapping algorithms in the estimation sample by study characteristics).

Figure 4. Comparison of Q-squared values between estimation sample, internal validation and external validation sample for studies that reported sufficient data to estimate Q-squared for at least two of these.



Abbreviations: OLS, Ordinary Least Squares; TPM, Two-part Model; MM, Mixture Model; RM, Response Mapping; CLAD, Censored Least Absolute Deviations.

Discussion

This study estimated Q-squared for mapping functions between HRQoL measures using studies identified from the Oxford HERC mapping database. Our study shows that it was straightforward to calculate Q-squared from routinely reported metrics in published mapping studies. Q-squared provided a useful way to compare mapping algorithms evaluated on different samples and using different measures. Q-squared allows prediction accuracy to be compared across samples, even when between-patient variability differs. This can be used for several purposes.

Firstly, it can be used, alongside other information, to choose which published mapping model to use to estimate target instrument scores in a dataset that did not directly measure the target instrument of interest: e.g., to estimate utilities for an economic evaluation. In this study, we were able to compare model performance for models mapping EORTC QLQ-C30 to EQ-5D, which suggested that the linear models by Kontodimopoulos et al.(20) produced the highest Q-squared in the estimation sample and were also among the best in external validation.(18) Having said that, we need to emphasize that other measures should also be considered alongside Q-squared, to ensure that the mapping algorithm does not systematically over- or under-estimate mean utility or differences between patient groups or give predictions outside the feasible range of the target instrument.(32) When choosing between mapping algorithms, it is also essential to consider the relevance of the data to the population of interest and the methodological quality of the study.

Secondly, the absolute Q-squared value can be used to benchmark the predictive value of a mapping model. We observed that Q-squared was numerically higher for algorithms mapping EORTC QLQ-C30 to EQ-5D-5L than those mapping from IBS-QoL and PAC-QOL. However,

our dataset focusing on GI disease studies is not sufficient to establish whether this reflects a weaker relationship between the source and target instruments, limitations in the data or the model specification used.

Q-squared can also be used to explore which subgroups of patients a mapping model performs better or worse in. In principle, Q-squared could also be used to weight models in model averaging(33); Q-squared may facilitate estimation of weights for response mapping or two-part models and allow weights to take account of the variability of the sample and the predictive power in the validation sample.

Within the small number of studies reporting Q-squared for the same algorithm in different samples, we did not observe Q-squared being higher in the estimation sample and lower in the external validation sample. Some studies (e.g. Sturkenboom et al. 2021)(30) were validated on very small samples. More research on a larger sample of studies is needed to explore this further.

To explore the usefulness of Q-squared in a manageable sample of studies, our study focused on a small number of studies in GI disease populations, which means the findings may not generalise to other diseases. The cut-offs that we used descriptively to define studies with substantial, moderate, weak and poor-performing models are based on those used for R^2 and have not been validated.(17) They covered the spread of Q-squared values for the mapping algorithms in our sample and may be helpful for other mapping studies. However, alternative cut-offs (e.g. (34,35)) may be more useful in other contexts where predictive models tend to have much lower accuracy.

In situations where the best mapping models have weak or poor predictive power, it may be particularly useful to review the conceptual overlap between the instruments and consider estimating a new mapping model on a larger dataset. Models with a negative Q-squared should not be used as they produce predictions that are worse than assuming that all patients have the mean value.

Future mapping studies should routinely report Q-squared alongside RMSE and MAE to facilitate comparisons between studies. Q-squared is simple to calculate for any predictive model and enables prediction accuracy to be directly compared between samples, which is not feasible with other metrics. It can be used alongside other information to inform the choice of mapping algorithm.

References

1. Devlin NJ, Brooks R. EQ-5D and the EuroQol Group: Past, Present and Future. *Appl Health Econ Health Policy*. 2017 Apr 1;15(2):127–37. doi:10.1007/s40258-017-0310-5
2. National Institute for Health and Care Excellence (NICE). Use of the EQ-5D-5L value set for England: NICE position statement. Published November 2, 2017. [Internet]. NICE; 2017 Nov [cited 2026 Feb 19]. Available from: <https://www.nice.org.uk/corporate/ecd16/chapter/introduction>
3. Devlin N, Pickard S, Busschbach J. The Development of the EQ-5D-5L and its Value Sets. In: Devlin N, Roudijk B, Ludwig K, editors. *Value Sets for EQ-5D-5L: A Compendium, Comparative Review & User Guide* [Internet]. Cham: Springer International Publishing; 2022 [cited 2026 Feb 19]. p. 1–12. Available from: https://doi.org/10.1007/978-3-030-89289-0_1 doi:10.1007/978-3-030-89289-0_1
4. Longworth L, Rowen D. Mapping to Obtain EQ-5D Utility Values for Use in NICE Health Technology Assessments. *Value in Health*. 2013 Jan 1;16(1):202–10. doi:10.1016/j.jval.2012.10.010
5. Dakin H, Burns R, Abel L, Koleva-Kolarova R, Yang Y, Mallon K, 2025. HERC database of mapping studies, Version 10.0 (Last updated: 7th November 2025). <https://doi.org/10.5287/ora-bdrvmy9mn>. Available at: <https://www.ndph.ox.ac.uk/research/research-groups/eph/resources/database-of-mapping-studies>.
6. Dakin H, Abel L, Burns R, Yang Y. Review and critical appraisal of studies mapping from quality of life or clinical measures to EQ-5D: an online database and application of the

MAPS statement. *Health and Quality of Life Outcomes*. 2018 Feb 12;16(1):31.

doi:10.1186/s12955-018-0857-3

7. Brazier JE, Yang Y, Tsuchiya A, Rowen DL. A review of studies mapping (or cross walking) non-preference based measures of health to generic preference-based measures. *Eur J Health Econ*. 2010 Apr 1;11(2):215–25. doi:10.1007/s10198-009-0168-z
8. Dakin HA, Gao N, Leal J, Holman RR, Tran-Duy A, Clarke P. Using QALYs as an Outcome for Assessing Global Prediction Accuracy in Diabetes Simulation Models. *Med Decis Making*. 2025 Jan;45(1):45–59. doi:10.1177/0272989X241285866 PubMed PMID: 39474832; PubMed Central PMCID: PMC11645849.
9. Petrou S, Rivero-Arias O, Dakin H, Longworth L, Oppe M, Froud R, et al. The MAPS Reporting Statement for Studies Mapping onto Generic Preference-Based Outcome Measures: Explanation and Elaboration. *Pharmacoeconomics*. 2015 Oct 1;33(10):993–1011. doi:10.1007/s40273-015-0312-9
10. Dakin H. Review of studies mapping from quality of life or clinical measures to EQ-5D: an online database. *Health and Quality of Life Outcomes*. 2013 Sep 5;11(1):151. doi:10.1186/1477-7525-11-151
11. Wailoo AJ, Hernandez-Alava M, Manca A, Mejia A, Ray J, Crawford B, et al. Mapping to Estimate Health-State Utility from Non–Preference-Based Outcome Measures: An ISPOR Good Practices for Outcomes Research Task Force Report. *Value in Health*. 2017 Jan 1;20(1):18–27. doi:10.1016/j.jval.2016.11.006

12. Quan NT. The Prediction Sum of Squares as a General Measure for Regression Diagnostics. *Journal of Business & Economic Statistics*. 1988;6(4):501–4. doi:10.2307/1391469
13. Wold H. Soft modelling: the basic design and some extensions. In: Joreskog K, Wold H, eds. *Systems under Indirect Observations: Causality, Structure, Predictions (Part 2)*. Amsterdam: North Holland; 1982. p 1–53.
14. Quan NT. The prediction sum of squares as a general measure for regression diagnostics. *J Bus Econ Stat*. 1988;6:501–4.
15. Villalva J. Structural equation models - PLS in engineering sciences: a brief guide for researchers through a case applied to the industry. *Athenea*. 2021 Jun 15;2:5–18. doi:10.47460/athenea.v2i4.17
16. Hair JF, Risher JJ, Sarstedt M, Ringle CM. When to use and how to report the results of PLS-SEM. *European Business Review*. 2019 Jan 14;31(1):2–24. doi:10.1108/EBR-11-2018-0203
17. Hair JF, Ringle CM, Sarstedt M. PLS-SEM: Indeed a Silver Bullet. *Journal of Marketing Theory and Practice*. 2011 Apr;19(2):139–52. doi:10.2753/MTP1069-6679190202
18. Woodcock F, Doble B, CANCER 2015 Consortium. Mapping the EORTC-QLQ-C30 to the EQ-5D-3L: An Assessment of Existing and Newly Developed Algorithms. *Med Decis Making*. 2018 Nov;38(8):954–67. doi:10.1177/0272989X18797588 PubMed PMID: 30226101.

19. Doble B, Lorgelly P. Mapping the EORTC QLQ-C30 onto the EQ-5D-3L: assessing the external validity of existing mapping algorithms. *Qual Life Res.* 2016 Apr 1;25(4):891–911. doi:10.1007/s11136-015-1116-2
20. Kontodimopoulos N, Aletras VH, Paliouras D, Niakas D. Mapping the cancer-specific EORTC QLQ-C30 to the preference-based EQ-5D, SF-6D, and 15D instruments. *Value Health.* 2009;12(8):1151–7. doi:10.1111/j.1524-4733.2009.00569.x PubMed PMID: 19558372.
21. Marriott ER, van Hazel G, Gibbs P, Hatswell AJ. Mapping EORTC-QLQ-C30 to EQ-5D-3L in patients with colorectal cancer. *J Med Econ.* 2017 Feb;20(2):193–9. doi:10.1080/13696998.2016.1241788 PubMed PMID: 27676291.
22. Franken MD, de Hond A, Degeling K, Punt CJA, Koopman M, Uyl-de Groot CA, et al. Evaluation of the performance of algorithms mapping EORTC QLQ-C30 onto the EQ-5D index in a metastatic colorectal cancer cost-effectiveness model. *Health and Quality of Life Outcomes.* 2020 Jul 20;18(1):240. doi:10.1186/s12955-020-01481-2
23. Kim SH, Jo MW, Kim HJ, Ahn JH. Mapping EORTC QLQ-C30 onto EQ-5D for the assessment of cancer patients. *Health and Quality of Life Outcomes.* 2012 Dec 17;10(1):151. doi:10.1186/1477-7525-10-151
24. Perwitasari DA, Purba FD, Candradewi SF, Marwin M, Permata A, Ulfa Faza MB, et al. Mapping EORTC-QLQ-C30 onto EQ-5D-5L Index in Indonesian Cancer Patients. *Asian Pac J Cancer Prev.* 2023 Apr 1;24(4):1125–30. doi:10.31557/apjcp.2023.24.4.1125 PubMed Central PMCID: PMC10352743.

25. Teckle P, McTaggart-Cowan H, Van der Hoek K, Chia S, Melosky B, Gelmon K, et al.
Mapping the FACT-G cancer-specific quality of life instrument to the EQ-5D and SF-6D.
Health and Quality of Life Outcomes. 2013 Dec 1;11(1):203. doi:10.1186/1477-7525-11-203
26. Yousefi M, Nahvijou A, Sari AA, Ameri H. Mapping QLQ-C30 Onto EQ-5D-5L and SF-6D-V2 in Patients With Colorectal and Breast Cancer From a Developing Country. Value in Health Regional Issues. 2021 May 1;24:57–66. doi:10.1016/j.vhri.2020.06.006
27. Hatswell AJ, Vegter S. Measuring quality of life in opioid-induced constipation: mapping EQ-5D-3 L and PAC-QOL. Health Economics Review. 2016 Apr 21;6(1):14.
doi:10.1186/s13561-016-0091-9
28. Parker M, Haycox A, Graves J. Estimating the relationship between preference-based generic utility instruments and disease-specific quality-of-life measures in severe chronic constipation: challenges in practice. Pharmacoeconomics. 2011 Aug;29(8):719–30.
doi:10.2165/11588360-000000000-00000 PubMed PMID: 21635017.
29. Fishman J, Higgins V, Piercy J, Pike J. Cross-walk of the Chronic Liver Disease Questionnaire for Nonalcoholic Steatohepatitis (CLDQ-NASH) and the EuroQol EQ-5D-5L in patients with NASH. Health Qual Life Outcomes. 2023 Oct 14;21(1):113.
doi:10.1186/s12955-023-02195-x PubMed Central PMCID: PMC10576276.
30. Sturkenboom R, Keszthelyi D, Brandts L, Weerts ZZRM, Snijkers JTW, Masclee AAM, et al. The estimation of a preference-based single index for the IBS-QoL by mapping to the EQ-5D-5L in patients with irritable bowel syndrome. Qual Life Res. 2022 Apr 1;31(4):1209–21. doi:10.1007/s11136-021-02995-y

31. Meunier A, Soare A, Chevrou-Severac H, Myren KJ, Murata T, Longworth L. Indirect and Direct Mapping of the Cancer-Specific EORTC QLQ-C30 onto EQ-5D-5L Utility Scores. *Appl Health Econ Health Policy*. 2022 Jan;20(1):119–31. doi:10.1007/s40258-021-00682-0 PubMed PMID: 34554442.
32. Hernández Alava M, Wailoo AJ, Ara R. Tails from the Peak District: Adjusted Limited Dependent Variable Mixture Models of EQ-5D Questionnaire Health State Utility Values. *Value in Health*. 2012 May 1;15(3):550–61. doi:10.1016/j.jval.2011.12.014
33. Jackson CH, Thompson SG, Sharples LD. Accounting for uncertainty in health economic decision models by using model averaging. *J R Stat Soc Ser A Stat Soc*. 2009 Apr;172(2):383–404. doi:10.1111/j.1467-985X.2008.00573.x PubMed PMID: 19381329; PubMed Central PMCID: PMC2667305.
34. Falk, R.F. and Miller, N.B. (1992) *A Primer for Soft Modeling*. University of Akron Press. - References - Scientific Research Publishing [Internet]. [cited 2026 May 10]. Available from: <https://www.scirp.org/reference/referencespapers?referenceid=2060943>
35. Chin WW. The partial least squares approach for structural equation modeling. In: *Modern methods for business research*. Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers; 1998. p. 295–336. (Methodology for business and management).

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and materials

All data supporting the findings of this study are available within the paper and its Supplementary Information.

Competing interests

None to declare

Funding

KM received funding from the Atlantic Technological University President's Bursary for a PhD scholarship. HD received funding from a Nuffield Department of Population Health Senior Fellowship.

Authors' contributions

All authors contributed to the study conception and design. Searches were performed by KM. KM conducted data analysis and interpretation with supervision from HD. KM and HD were involved in manuscript drafting and revision, and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Acknowledgements

Not applicable

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryTable1.docx](#)
- [SupplementaryTable2Dataextractiontable.xlsx](#)
- [SupplementaryTable3.docx](#)