

Systematic Literature Review

Statistical Methods for Analyzing EQ-5D in Randomized Clinical Trials: A Systematic Literature Review

Jiajun Yan, MSc, Brittany Humphries, PhD, Ruinan Xie, MSc, Ziran Yin, MSc, Zhenyan Bo, MSc, Sha Diao, MSc, Jing Cai, MSc, Preston Tse, BHSc, Meixuan Li, PhD, Eleanor Pullenayegum, PhD, Shun Fu Lee, PhD, Feng Xie, PhD

ABSTRACT

Objectives: We conducted a systematic literature review to summarize the application of statistical methods for analyzing treatment effect on EQ-5D in randomized clinical trials (RCTs).

Method: We searched 2 electronic databases (MEDLINE and EMBASE, from inception through 2021) and www.clinicaltrial.gov. Eligible studies were RCTs that analyzed postbaseline EQ-5D data by treatment group. Information on trial characteristics, EQ-5D data characteristics, and statistical methods were extracted. Descriptive statistics were used to summarize results by dimension response, EQ visual analog scale (EQ VAS), and EQ-5D utility.

Results: A total of 2125 trials met the eligibility criteria. EQ-5D was commonly considered a secondary ($n = 1219$, 57.4%) or exploratory ($n = 775$, 36.5%) endpoint in RCTs. EQ-5D utilities were the most analyzed. Both utilities and EQ VAS were primarily analyzed in numerical format. The most common statistical models for analyzing utilities were the linear fixed-effect model for single postbaseline (192/589, 32.6%) and the linear mixed-effect model for multiple post-baselines (338/984, 34.3%). Of the 2054 studies that analyzed numerical EQ-5D, 221 (10.8%) examined model assumptions and 438 (21.3%) adjusted for the baseline score. Missing data were explicitly assessed in 661 trials, among which 347 (52.5% of 661) applied imputations, with the 2 most used imputation methods being multiple imputations ($n = 200$, 57.6% of 347) and last observation carried forward ($n = 106$, 30.5% of 347).

Conclusions: This review found that health utilities are the most frequently analyzed EQ-5D data collected in clinical trials, followed by EQ VAS. Significant variation was observed in the selection of models, with most trials lacking adjustments for baseline data and appropriate methods for handling missing data.

Keywords: EQ-5D, patient-reported outcome, randomized clinical trial, statistical analysis.

VALUE HEALTH. 2025; 28(7):1126–1135

Highlights

- The EQ-5D, one of the most widely used preference-weighted, patient-reported outcome measures, has been commonly collected in randomized clinical trials. However, it is not clear how it has been analyzed and reported.
- A systematic review of published literature and clinical trial registry reveals a significant variation in the models chosen for analyzing EQ-5D data in trials, along with issues such as lack of baseline adjustment and missing data handling.
- Given the growing interest in using preference-weighted instruments in clinical trials, this review provides insight into the application of statistical methods and serves as a basis for developing statistical analysis guidelines for this type of data, enhancing their value in informing clinical and economic evaluations.

Introduction

A randomized clinical trial (RCT) is a prospective study design that randomly assigns participants to an intervention or control group. RCT is considered to be the gold standard for evaluating the efficacy and safety of a new treatment.¹ Although clinical metrics are commonly used as primary endpoints in RCTs, patient-reported outcomes (PROs) have been increasingly recognized as an important outcome by regulatory authorities, reimbursement agencies, guideline developers, and policy makers worldwide.^{2–7}

The EQ-5D is one of the most used generic, preference-based PRO instruments. It has a descriptive system and a visual analog scale (EQ VAS). The descriptive system includes 5 dimensions: mobility, self-care, usual activities, pain/discomfort, and anxiety/depression.^{8–12} The original version of the EQ-5D (ie, EQ-5D-3L) has 3 response options for each dimension, indicating “no” (usually coded as 1), “some” (2), and “extreme problems” (3). The responses to the descriptive system can be described as a 5-digit

string, with 11111 being the best health state (no problem in any dimension) and 33333 the worst (extreme problems in all dimensions). The EQ-5D-3L describes a total of 243 (=3⁵) health states. The 3-level response options were later expanded into 5 levels of severity (ie, EQ-5D-5L): no, slight, moderate, severe, and extreme, which describes 3125 (=5⁵) health states.^{12,13} A health state can be converted into a health utility value by applying a scoring algorithm that is based on preferences obtained from the general population. Health utility values anchor at 1 for full or perfect health and 0 for being dead. Negative values can be assigned to health states considered worse than dead.^{14–19} The EQ VAS measures global health status on a vertical 20 cm scale ranging from 0 (worst imaginable health on the bottom of the scale) to 100 (best imaginable health on the top of the scale).^{12,13} The dimension response, EQ-5D utility, and the EQ VAS are the 3 distinct types of data generated from the EQ-5D.²⁰

As a generic preference-based measure of health-related quality of life (HRQoL), the EQ-5D has found broad applications

in clinical trials, economic evaluations, population health surveys, and routine clinical practice.^{21–24} There is a guidance book on analyzing and reporting EQ-5D data, but it mostly focuses on analysis for cross-sectional data or single cohort. In section 4.7, the authors briefly list a few regression models available for analyzing treatment effects using the EQ-5D but without any further details.²⁰ It is not clear what and how statistical methods have been applied to analyze EQ-5D data collected in RCTs. Therefore, we conducted this systematic literature review to summarize the statistical methods that have been used to analyze treatment effects on HRQoL as measured using the EQ-5D within the context of RCTs.

Methods

Trial Identification

A systematic literature search of 2 electronic databases (EMBASE and MEDLINE) was conducted from their inception through November 15, 2021. The search strategy combined key terms related to the study design (eg, RCT) and the terms used to describe the EQ-5D. (see [Appendix Table 1](#) in [Supplemental Materials](#) found at <https://doi.org/10.1016/j.jval.2025.02.001>) In addition, we also used the term “EQ-5D” and searched www.clinicaltrial.gov, a registry database for clinical trials, on August 16, 2023. For every record identified, we recorded the National Clinical Trial number and searched for corresponding publications in PubMed. If no publication could be found on PubMed, we reviewed and extracted the methods and results as posted on the registry website.

Trial Eligibility

We included RCTs that analyzed postbaseline EQ-5D data by treatment group. There were no restrictions for the patient population or type of treatment received. Outcomes of interest included EQ-5D dimension responses, EQ-5D utilities, or EQ VAS. Given that the review focused on estimating treatment effect on EQ-5D, we did not include other EQ-5D-derived outcomes, such as quality-adjusted life years or quality-adjusted time without symptoms of disease or toxicity. These outcomes are composite measures that incorporate both time and QoL. Analyses of these outcomes differ from analyses of health utilities in terms of data considerations and analytical methods. This exclusion also avoids a potentially large amount of literature on economic evaluation as a secondary use of the trial data. Peer-reviewed journal articles and health technology assessment (HTA) reports were eligible for inclusion. We excluded pilot and feasibility studies, reviews, editorials, conference abstracts, and non-English language reports.

Trial Selection

Title and abstract screening were performed independently and in duplicate by 8 reviewers (J.Y., R.X., Z.Y., Z.B., S.D., J.C., P.T., and M.L.) using Covidence.²⁵ Subsequent full-text screening was done independently by pairs of reviewers using Microsoft Excel. Any discrepancies between reviewers were resolved by discussion with a third reviewer. To prevent double counting, studies with multiple publications (eg, a primary clinical article, a secondary PRO analysis article, or an HTA report) were mapped to trial ID by using the registration number, indication, treatments, and sample size.

Data Extraction and Categorization

Data were extracted independently by 4 pairs of reviewers using a standardized template in Microsoft Excel that was

developed and pilot tested by the research team. The following information was extracted from the included studies:

1. Trial Characteristics: trial phase, starting year, design, therapeutic area, sample size, and follow-up duration.
2. EQ-5D data characteristics: type of study outcome for the EQ-5D (ie, primary or secondary outcome. If the study did not specifically identify EQ-5D as a primary or secondary outcome, it was classified under exploratory outcomes. This includes instances in which the EQ-5D was identified as a PRO or HRQoL outcome.), EQ-5D data collection time points, EQ-5D data analyzed (dimension response, EQ-5D utility, or EQ VAS), and variable format (numerical and categorical). For the variable format, EQ-5D utilities and EQ VAS scores are numerical data. They can also be transformed into categorical data for analysis. For example, a minimal important difference (MID) can be used to classify the data into categories such as “deterioration or no deterioration” and “improvement or no improvement” (eg, a change in utility $>$ MID vs a change in utility $\leq$ MID); or a threshold can be set to define health status categories such as “good” (eg, utility $>$ 0.8) or “bad” health. Dimension responses contain 5 multilevel (3 or 5) categorical responses. These can be simplified into binary data by grouping the responses into “no problem” (level 1) vs “any problem” (level 2 and higher). Numerical values could also be assigned to the levels of categorical responses (ie, 1 to level 1, 2 to level 2, and so forth), which are then treated as numerical variables.
3. Statistical methods: methods used to estimate the treatment effect on the EQ-5D, assumptions examined (eg, normal distribution and homoscedasticity), and methods for adjusting for confounding factors and handling missing data.

We broadly grouped the statistical methods into descriptive, bivariate, and multivariable methods. The descriptive methods focus on the description of a single variable (ie, EQ-5D data). The bivariate methods involve the EQ-5D with another variable (ie, treatment arms), which may include parametric (eg, t test or analysis of variance) and nonparametric tests (eg, Mann-Whitney U test, χ^2 test) or simple regression analysis. The multivariable methods include regression models involving the EQ-5D and multiple variables. Linear or logistic regressions were further categorized into fixed-effect models and mixed-effect models. Generalized estimating equations were another longitudinal multivariable analysis that can be used to handle both numerical and categorical variables.^{26–28} We listed survival analyses as an independent category because these methods focus on time to event (ie, time to deterioration) or hazard instead of the magnitude of difference in EQ-5D or proportion of improvement.

Data Synthesis

We summarized the results by type (ie, dimension response, EQ VAS, and EQ-5D utility), and format (ie, numerical and categorical) of EQ-5D data analyzed. We also reported the analyses separated for single postbaseline follow-up and multiple follow-ups. This is because subjects’ intermediate measures are correlated with their prior data, yielding a within-subject correlation, which can be expressed in a matrix. In the case of multiple follow-ups, this matrix becomes high dimensional and requires additional consideration in the analysis. In contrast, a single follow-up design results in a simpler correlation between baseline and follow-up, which can be addressed using paired data analysis (eg, analyzing change from baseline) without the need for a correlation matrix. R statistical software (version 4.3.3)²⁹ was used for data analysis and visualization. The results are reported according

to Preferred Reporting Items for Systematic Reviews and Meta-Analyses guidelines.³⁰

Quality Appraisal

The objective of this review was to summarize statistical methodology and not to estimate any treatment effect. As a result, we did not perform any quality appraisal on the included RCTs using the Cochrane risk-of-bias tool³¹ or Grading of Recommendations, Assessment, Development, and Evaluations.³²

Results

Summary of Study Selection

The Preferred Reporting Items for Systematic Reviews and Meta-Analyses flow diagram is presented in Figure 1. A total of 11 633 records were identified in the searches of electronic databases and clinicaltrials.gov, of which 9030 abstracts and titles were screened, and 7056 were included in the full-text screening. The final analysis included 2125 unique RCTs.

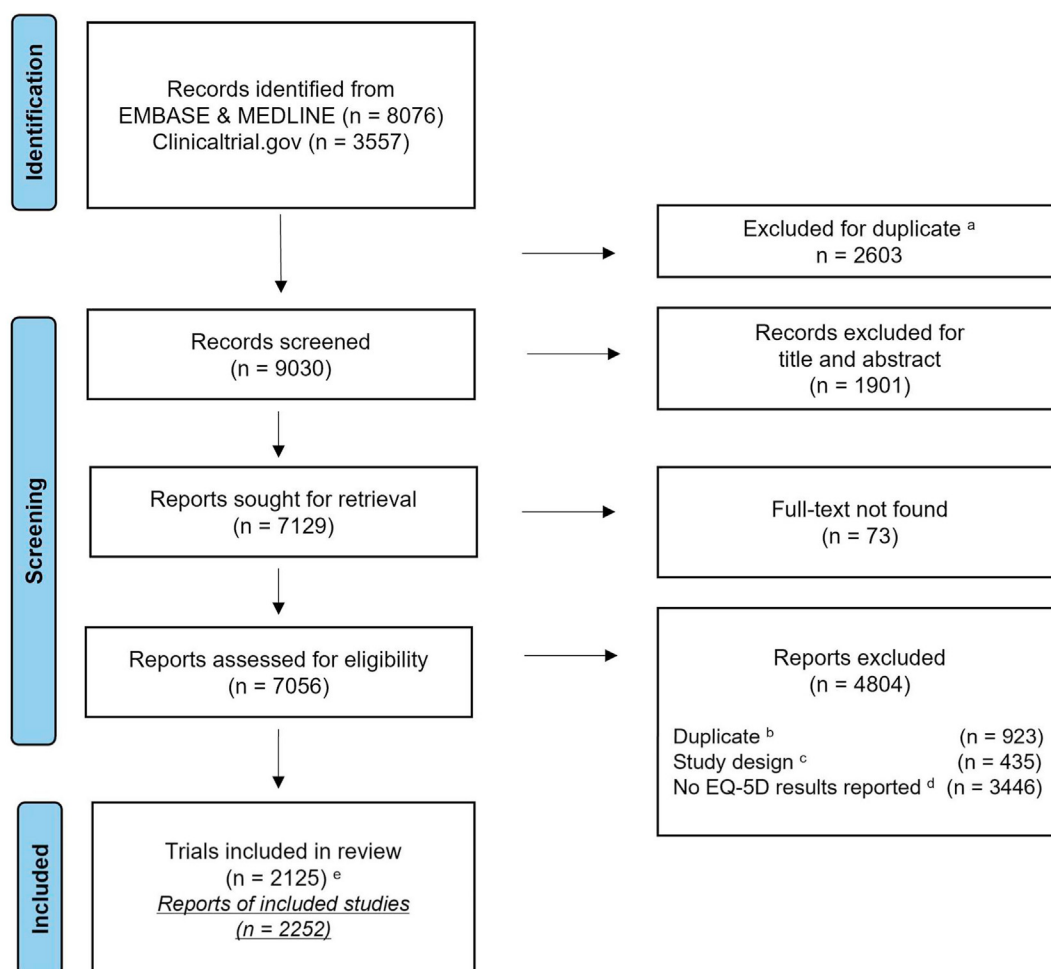
Characteristics of the Trials

Table 1 presents the characteristics of the 2125 RCTs. A total of 1316 (61.9%) were nondrug trials (eg, of surgical procedures, devices, or diagnostic tests) that did not specify the trial phase. Among the remaining 809 (38.1%) trials, 574 were phase 3 trials. Parallel treatment design was used in 2044 trials (96.2%). The top-5 therapeutic areas were orthopedics ($n = 349$, 16.4%), cardiology ($n = 296$, 13.9%), oncology ($n = 265$, 12.5%), psychiatry ($n = 172$, 8.1%), and rheumatology ($n = 161$, 7.6%). Only 131 (6.2%) trials measured the EQ-5D as a primary outcome. Instead, EQ-5D was more frequently measured as a secondary ($n = 1219$, 57.4%) or exploratory ($n = 775$, 36.5%) outcome. Baseline EQ-5D was collected in 1881 trials (88.5%). 855 trials (40.2%) had single postbaseline data collection and 1270 (59.8%) had more than 1 postbaseline data collection.

EQ-5D Data Characteristics

EQ-5D utilities were the most frequently analyzed type of EQ-5D data ($n = 1592$, 74.9%). EQ VAS and dimension responses were analyzed in 1197 (56.3%) and 385 (18.1%) trials, respectively.

Figure 1. The preferred reporting items for systematic reviews and meta-analyses flow diagram.



^aAutomatic de-duplication by EndNote 20 and Covidence. ^bDue to the random assignments of abstract screening tasks in Covidence, duplicated abstracts might not be reviewed by one person. Therefore, study de-duplication ($n=923$) was pushed to full-text screening stage. ^cNot RCTs, e.g., pilot studies, feasibility studies, reviews, etc. ^dStudy protocols were also excluded for "No EQ-5D results posted". ^eStudies with multiple publications (e.g., a primary clinical paper, a secondary PRO analysis paper, or a HTA report) were mapped to trial ID by using the registration number, indication, treatments, and sample size.

Table 1. Characteristics of included trials.

Study characteristics	Number of studies, n (%) (N = 2125)
Trial phase	
Phase	809 (38.1%)
1	8 (1.0%)
2	121 (5.0%)
3	574 (71.0%)
4	106 (13.1%)
Unspecified*	1316 (61.9%)
Treatment arm design	
Parallel group	2044 (96.2%)
Crossover	51 (2.4%)
Factorial	28 (1.3%)
Others	2 (0.1%)
Therapeutic area	
Orthopedics	349 (16.4%)
Cardiology	296 (13.9%)
Oncology	265 (12.5%)
Psychiatry	172 (8.1%)
Rheumatology	161 (7.6%)
Neurology	137 (6.4%)
Endocrinology	105 (4.9%)
Dermatology	76 (3.6%)
Primary care	75 (3.5%)
Pulmonology	66 (3.1%)
Others†	423 (19.9%)
Type of endpoint for EQ-5D	
Primary	131 (6.2%)
Secondary	1219 (57.4%)
Exploratory‡	775 (36.5%)
Baseline EQ-5D collection	
Yes	1881 (88.5%)
No	202 (9.5%)
Unknown	42 (2%)
# of postbaseline EQ-5D collection	
1	855 (40.2%)
≥2	1270 (59.8%)

*Nondrug trials (on medical devices and surgical or diagnostic procedures) that did not specify the trial phase.

†Others included analgesia/anesthesiology/anti-inflammatory, critical care, dietetics, gastroenterology, geriatrics, gynecology, hematology, hepatology, hospital management, immunology, infectious diseases, multimorbidity, nephrology, obstetrics, ophthalmology, otolaryngology, pain management, pediatrics, psychology, radiology, sleep management, and urology.

‡Trials that did not specify EQ-5D as a primary or secondary outcome were classified under exploratory outcomes, eg, EQ-5D was a patient-reported outcome or a health-related quality-of-life outcome.

Table 2 shows the contingency table of the 3 EQ-5D data analyzed. Most trials considered utility only ($n = 777$, 36.6%) or utility and EQ VAS ($n = 661$, 31.1%). Utility and dimension response ($n = 55$, 2.6%) and dimension response only ($n = 88$, 4.1%) were the least analyzed.

Both utilities and EQ VAS were analyzed predominantly in their original numerical format ($n = 1573$, 98.8% of 1592 and $n = 1186$, 99.1% of 1197, respectively) (Fig. 2). As shown in Figure 2A, EQ-5D utilities were analyzed categorically by defining health status as good/bad health ($n = 6$, 0.4% of 1592) or as stable/deterioration/improvement ($n = 51$, 3.2% of 1592). Of these, 41 trials categorized the utilities using MID. Figure 2B shows that EQ VAS were categorized as good/bad health ($n = 4$, 0.3% of 1197) or stable/deterioration/improvement ($n = 40$, 3.3% of 1197), with 32 trials applying MID for the categorization.

Table 2. Frequency of EQ-5D data analyzed in randomized clinical trials.

Utility	EQ VAS	Dimension response	# of RCTs, n (%)
N = 1592	N = 1197	N = 385	N = 2125*
Yes	No	No	777 (36.6%)
Yes	Yes	No	661 (31.1%)
No	Yes	No	294 (13.8%)
No	Yes	Yes	143 (6.7%)
Yes	Yes	Yes	99 (4.7%)
No	No	Yes	88 (4.1%)
Yes	No	Yes	55 (2.6%)

EQ VAS indicates EQ visual analog scale; RCT, randomized clinical trial.

*Eight studies reported EQ-5D but did not specify the type of EQ-5D data.

As shown in Figure 2C, EQ-5D dimension responses were analyzed in categorical ($n = 231$, 60.0% of 385) or numerical format ($n = 156$, 40.5% of 385). Most categorical analyses used the percent of dimension responses ($n = 193$, 50.1% of 385).

Statistical Methods Used to Analyze EQ-5D

Numerical format

Figure 3 presents a summary of the statistical methods used to analyze numerical EQ-5D data. Among 589 trials that analyzed single postbaseline utility, 84 (14.3%) used descriptive statistics, 221 (37.6%) bivariate methods, and 273 (46.3%) multivariable. The most frequently used method was linear fixed-effect model ($n = 192$, 32.6%), followed by parametric test ($n = 121$, 20.5%) and nonparametric test ($n = 83$, 13.9%). Among 984 trials that conducted analyses on multiple postbaseline utilities, 110 (11.2%), 299 (30.4%), and 612 (62.6%) used descriptive, bivariate, and multivariable methods, respectively. Linear mixed-effect model ($n = 338$, 34.3%) was the most frequently used, followed by the parametric test ($n = 220$, 22.4%) and fixed-effect model ($n = 196$, 19.9%).

For EQ VAS, among 489 trials with single postbaseline, descriptive, bivariate, and multivariable methods were used in 89 (18.2%), 201 (41.1%), and 192 (39.1%) trials, respectively. Fixed-effect model ($n = 137$, 28.0%), parametric test ($n = 108$, 22.1%), and nonparametric test ($n = 71$, 14.5%) were among the most frequently used. Among 697 trials with multiple postbaseline measurements, descriptive, bivariate, and multivariable methods were used in 93 (13.3%), 216 (31.0%), and 427 (61.3%) trials, respectively, with mixed-effect model ($n = 245$, 35.2%), parametric test ($n = 129$, 18.5%), and fixed-effect model ($n = 121$, 17.4%) being the most frequently used.

Only 72 trials with single postbaseline and 84 trials with multiple post-baselines analyzed EQ dimension responses in numerical format. The most used methods for analyzing numerical dimension scores included parametric test, nonparametric test, fixed-effect model, and mixed-effect model, which aligned with utility and EQ VAS analysis methods (Fig. 3).

Other models, including Tobit regression, beta regression, Poisson regression, and quantile regression, were seen in very few studies (<3 for each type of EQ-5D).

Of the 2054 studies that analyzed numerical EQ-5D data, 221 (10.8%) reported that they examined method assumptions and 438 (21.3%) adjusted for the baseline score as a covariate. Only 54 (2.6%) trials used MID to interpret the clinical meanings of the treatment effect estimates.

Categorical format

The statistical methods applied to analyze categorical EQ-5D data are shown in Fig. 4. Nonparametric test (ie, chi-square test or Fisher's exact test) was frequently used across all types of EQ-5D data (>25.0% in either single or multiple postbaseline analyses). Descriptive analysis was also regularly used in single postbaseline analyses for all EQ-5D data (>20.0%) and multiple postbaseline analyses for dimension responses (47/114, 41.2%). Other common methods (>20.0% in any data analyses) included fixed-effect logistic model and Cochran-Mantel-Haenszel test.

Survival Analysis

Seventeen trials applied survival analysis to estimate the time to deterioration in utility, whereas 18 trials applied it to EQ VAS (see Appendix Table 3 in Supplemental Materials found at <https://doi.org/10.1016/j.jval.2025.02.001>)

Missing Data

Missing EQ-5D data were explicitly examined in 661 (31.1%) trials, among which 347 (52.5% of 661) conducted imputations. The most common imputation methods were multiple imputation

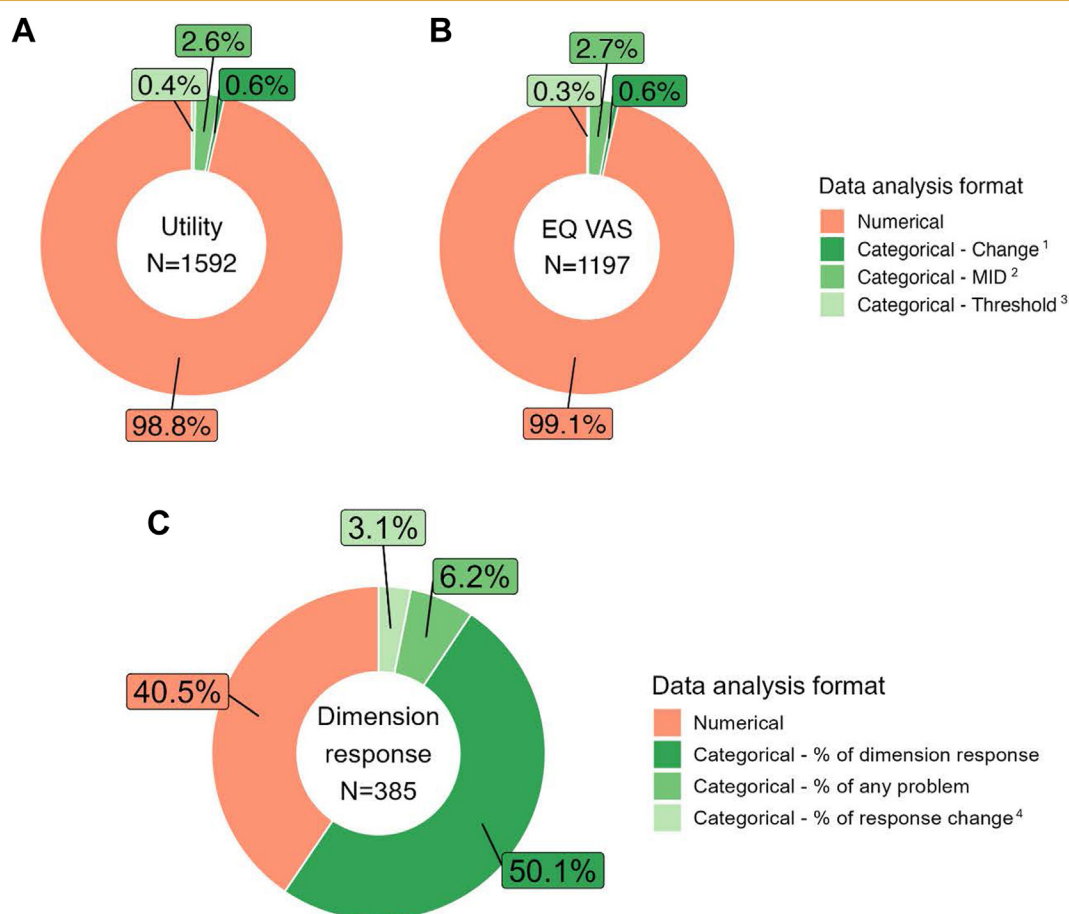
($n = 200$), and the last observation carried forward ($n = 106$). Others included imputations with mean, regression, Markov Chain Monte Carlo, nonresponder, and worst case.

Discussion

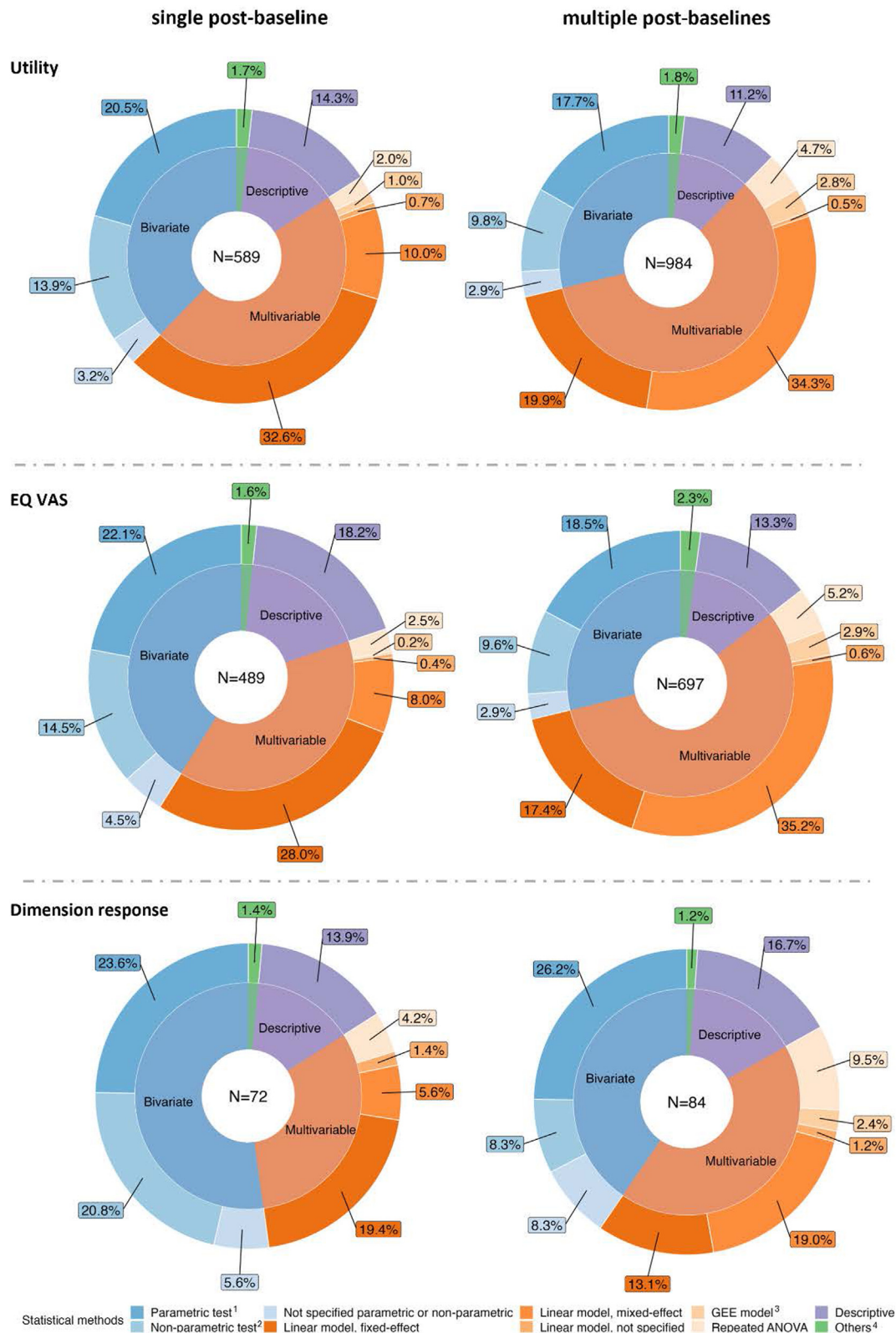
This systematic review summarized the statistical methods used to analyze EQ-5D data collected in RCTs. The EQ-5D utilities are the most frequently analyzed data. A wide range of descriptive, bivariate, and multivariable statistical methods have been used on EQ-5D data. The most frequently used models were the fixed-effect model for single postbaseline data and the mixed-effect model for multiple postbaseline data. However, no single method emerged as dominant across studies.

Among the 3 types of EQ-5D data, utilities and EQ VAS were primarily used to analyze treatment effects in the RCTs, whereas the dimension responses were rarely reported or analyzed. A similar usage pattern was also seen in HTA and regulatory claims.³³ As numerical summary measures, utilities and EQ VAS offer certain advantages for analysis, such as conventional statistical modeling and intuitive interpretation. To further facilitate

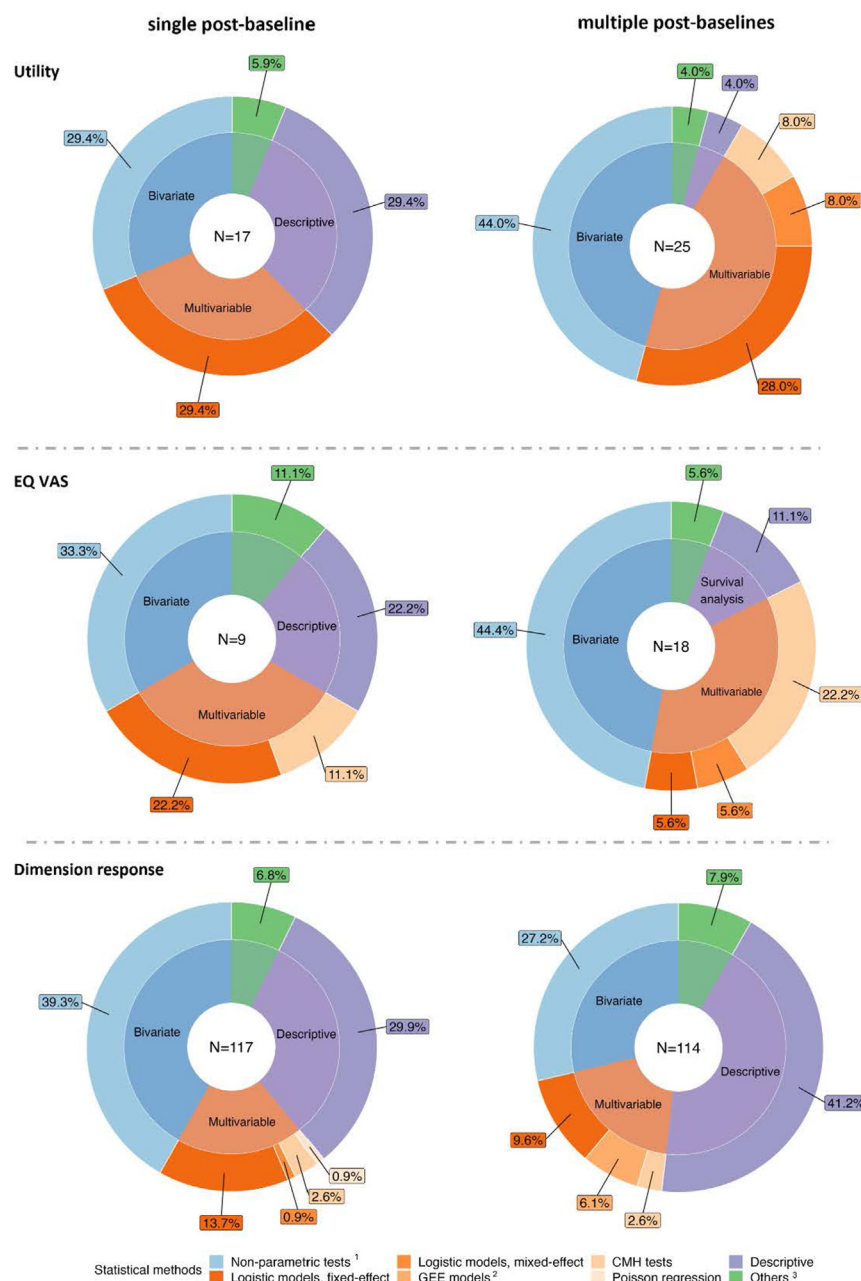
Figure 2. Variable format of utility, EQ VAS, and dimension response. (A) Proportion of variable formats used for utility, (B) proportion of variable formats used for EQ VAS, and (C) proportion of variable formats used for dimension response.



The sum of percentages may exceed 100% due to multiple formats used in trials. 16 studies were not included: 8 reported EQ-5D but did not describe the type of the data and 8 (1 on utility and 7 on dimension response) did not describe variable format. ¹Categorized as improved/deteriorated/stable/ by positive/negative/no change from baseline. ²Categorized as improved/deteriorated/stable/ by comparing change from baseline with minimal important difference (MID). ³Categorized as full/good and bad health by specific thresholds, eg, utility > 0.78 was considered as good health while < 0.78 was bad health. Thresholds for utility included 1, 0.78, 0.6, and 0.5 and population norms while thresholds for EQ VAS included 100 and 60. ⁴Response change included improved/stable/deteriorated/mixed change.

Figure 3. Statistical methods for numerical utility, EQ VAS, and dimension response analyses by single versus multiple post-baselines.

The sum of percentages may exceed 100% since multiple methods could be applied in one study. Unknown methods were not included. Details can be found in Appendix Table 2 in Supplemental Materials found at <https://doi.org/10.1016/j.jval.2025.02.001>. ¹Parametric tests included t-test and ANOVA. ²Non-parametric tests included Mann-Whitney U test and Kruskal-Wallis test. ³GEE models: Generalized estimating equations models. ⁴Others included beta regression models, constrained longitudinal data analysis, cumulative probability models, general additive mixed model, growth curve model, median regression model, pattern mixture mixed model, Poisson regression model, quantile regression model, regression with natural spline function, tobit regression model, two-part models, and weighted clustered regression models.

Figure 4. Statistical methods for categorical utility, EQ VAS, and dimension response analyses by single versus multiple post-baselines.

The sum of percentages may exceed 100% since multiple methods could be applied in one study. Unknown methods were not presented in this pie chart. Details can be found in [Appendix Table 3](#) in [Supplemental Materials](#) found at <https://doi.org/10.1016/j.jval.2025.02.001>. ¹Non-parametric tests included chi-square test and Fisher's exact test. ²GEE models: Generalized estimating equations models. ³Others included ANOVA, Kruskal-Wallis test, linear regressions, Mann-Whitney U test, ordered polytomous regression for repeated measures, and t-test. ANOVA, Kruskal-Wallis test, linear regressions, Mann-Whitney U test, and t-test were categorized in "others" category because they were applied to the values of percentages, which were not considered as categorical variables.

interpretation, MID's have been developed for EQ-5D utility in various disease areas^{34,35} (eg, cancer,³⁶ stroke,³⁷ rheumatoid arthritis,³⁸ and other chronic health conditions³⁹), thus increasing the usage of utilities. However, utility is not simply a single number but a combination of both dimension responses and value sets. The choice of value sets might influence the utilities because different value sets can yield varying utility values, potentially affecting treatment effect estimates. In contrast, dimension

responses can provide important nuances specific to each dimension. For example, Hagiwara et al⁴⁰ and Nauck et al⁴¹ found statistically significant treatment effects on overall utility, whereas certain dimensions, such as anxiety/depression, showed no significant effects. To enhance the analysis of dimension responses, the Paretian Classification of Health Change has been introduced,²⁰ although we observed very few applications in practice.⁴²⁻⁴⁴

EQ-5D utilities and EQ VAS are known to have special distributional characteristics, such as skewness, ceiling effects, and discreteness, which pose challenges to the assumptions required by regression models, eg, the normality assumption of linear regression.^{45,46} In this SLR, we observed the frequent use of linear regressions. However, few studies examined these assumptions. Violation of model assumptions could lead to biased or inefficient estimates of treatment effects. Moreover, this issue, combined with the use of diverse methods across studies, may reduce the comparability of results that complicate the evidence synthesis in systematic reviews and meta-analyses. Other methods, such as the Tobit model,⁴⁷⁻⁵¹ beta model,^{52,53} generalized additive models,⁵⁴ and adjusted limited dependent variable mixture model,^{55,56} can theoretically handle EQ-5D's unique characteristics better. However, each model also has its own strengths and limitations, and it remains unclear how different these models are in practice. Furthermore, variation in the models chosen highlights the lack of consensus on best practices in the field and indicates the need for further research to compare different models.

Simulation studies and guidelines have recommended controlling for potential baseline score differences in analyses, which will increase the validity and reliability of estimates.^{34,57-63} However, less than 30% of RCTs explicitly reported adjusting for the baseline EQ-5D score in their analyses. In addition, missing data were explicitly examined in only 31.1% (661/2125) of included trials. Mixed models have been recommended by guidelines^{64,65} because they can handle missing data by implicitly imputing the missingness under the missing-at-random assumption.^{66,67} In the reviews of PRO analysis methods, Pe et al⁶⁸ and Qian et al⁶⁹ found an increasing trend of using mixed models in recent years. However, when introducing mixed models for longitudinal analysis, Cnaan et al⁷⁰ expressed concern regarding the mechanism for missingness in quality-of-life measure. Patients with poor quality of life (eg, due to adverse events or lack of treatment efficacy) are less likely to contribute to the self-reported questionnaires, which potentially violates the missing-at-random assumption in the mixed models.⁷⁰ To evaluate the robustness of analysis results under different missing data assumptions, Scharfstein et al⁷¹ developed a semi-parametric method for global sensitivity analysis in clinical trials and demonstrated its application through a clinical trial with missing PRO.⁶⁵

The choice of analysis methods should be guided by the study hypothesis or the estimands—the true treatment effects that the study aims to estimate.^{72,73} For example, in a longitudinal trial, if the change in an outcome over time is an estimand, then it would be more appropriate to apply a mixed model which includes all time points simultaneously, incorporating within-subject correlation. Pe et al⁶⁸ found few trials reported hypotheses in their PRO analyses, and recent research and guidelines highlighted the need for improvement in the description of estimands for clinical trials.^{68,74-78}

Our review has a few strengths. This is the first systematic review that explores statistical methods specifically applied for EQ-5D data analysis. Considering the characteristics of EQ-5D, we described our findings based on the types and formats of EQ-5D data. Because the EQ-5D is often a secondary or exploratory outcome in RCTs, it may not be reported in titles and abstracts or included as a keyword in peer-reviewed journal articles. This may affect the sensitivity of electronic database search. In addition, a nonsignificant result for an exploratory outcome may discourage reporting, potentially leading to publication bias. We additionally searched the clinical trial registry database to ensure our review is comprehensive. Another strength is that we applied statistical knowledge to address variations in reporting statistical methods. For example, linear mixed model, also known as a multilevel

model or hierarchical model, was grouped as a linear mixed-effect model in our analysis; analysis of covariance and multiple linear regression were categorized as linear fixed-effect models.

Our review has a few limitations. We did not impose a time constraint on the literature search. The usage of methods for PRO analyses has changed in recent years compared with 20 years ago.^{68,69} However, we found that the difference was small. Commonly used methods (eg, linear regressions and *t* test) have remained popular through decades. A wide time window allowed us to gain a comprehensive understanding of the methods that have been widely used. Another limitation is that most publications were clinical trials with limited space for descriptions of the statistical methods and EQ-5D results. Therefore, the reporting of nonprimary outcomes was even more sparse, which affected the completeness of the information we extracted. For example, many studies did not provide details about their models (eg, covariates and correlation matrix); thus, the number of trials that accounted for baseline adjustment or applied missing data imputation may be underestimated. Other authors have recognized this issue and called for standardized reporting of PRO data in oncology RCTs.⁶⁸

Conclusion

This review found that health utilities are the most frequently analyzed EQ-5D data collected in clinical trials, followed by EQ VAS. Despite describing health profiles, dimension responses were only analyzed in a small proportion of the trials. Significant variation was observed in the selection of regression models, with most trials lacking adjustments for baseline data and appropriate methods for handling missing data.

Author Disclosures

Author disclosure forms can be accessed below in the [Supplemental Material](#) section.

Supplemental Material

Supplementary data associated with this article can be found in the online version at <https://doi.org/10.1016/j.jval.2025.02.001>.

Article and Author Information

Accepted for Publication: February 4, 2025

Published Online: March 17, 2025

doi: <https://doi.org/10.1016/j.jval.2025.02.001>

Author Affiliations: Department of Health Research Methods, Evidence, and Impact, McMaster University, Hamilton, ON, Canada (Yan, Humphries, Tse, Lee, F. Xie); Department of Applied Mathematics, Hong Kong Polytechnic University, Hong Kong, China (R. Xie); Department of Biostatistics and Bioinformatics, School of Medicine, Duke University, Chapel Hill, NC, USA (Yin); Department of Pharmacy/Evidence-based Pharmacy Center, West China Second University Hospital, Sichuan University, Sichuan, China (Bo, Diao); Department of Mathematics and Statistics, McMaster University, Hamilton, ON, Canada (Cai); Evidence-Based Medicine Center, School of Basic Medical Sciences, Lanzhou University, Lanzhou, China (Li); Health Technology Assessment Centre, Evidence-Based Social Science Research Center, School of Public Health, Lanzhou University, Lanzhou, China (Li); Child Health Evaluative Sciences, Hospital for Sick Children, Toronto, ON, Canada (Pullenayegum); Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada (Pullenayegum); Population Health Research Institute, Hamilton Health Sciences, Hamilton, ON, Canada

(Lee); Centre for Health Economics and Policy Analysis, McMaster University, Hamilton, ON, Canada (F. Xie).

Correspondence: Feng Xie, PhD, Department of Health Research Methods, Evidence and Impact, McMaster University, 1280 Main St West, Hamilton, ON L8S 4K1, Canada. Email: fengxie@mcmaster.ca

Authorship Confirmation: All authors certify that they meet the ICMJE criteria for authorship.

Funding/Support: This study was funded by a PhD grant from the EuroQol Research Foundation (# 345-PHD).

Role of Funder /Sponsor: The funding body had no role in the design of the study, the collection, analysis, and interpretation of data, the writing of the manuscript, or in the decision to submit the article for publication.

REFERENCES

- Hariton E, Locascio JJ. Randomised controlled trials - the gold standard for effectiveness research: study design: randomised controlled trials. *BJOG*. 2018;125(13):1716.
- Department of Health and Human Services FDA Center for Drug Evaluation and Research; U.S. Department of Health and Human Services FDA Center for Biologics Evaluation and Research; U.S. Department of Health and Human Services FDA Center for Devices and Radiological Health. Guidance for industry patient-reported outcome measures: use in medical product development to support labeling claims: draft guidance. *Health Qual Life Outcomes*. 2006;4:79.
- European Medicines Agency. The guideline on the evaluation of anticancer medicinal products in man. Appendix 2: The use of patient-reported outcome (PRO) measures in oncology studies (EMA/CHMP/292464/2014). [PDF]. https://www.tga.gov.au/sites/default/files/2024-05/EMACHMP292464_2014%20appendix-2-guideline-evaluation-anticancer-medicinal-products-man_en.PDF; 2016. Accessed March 12, 2025.
- Teixeira MM, Borges FC, Ferreira PS, Rocha J, Sepodes B, Torre C. A review of patient-reported outcomes used for regulatory approval of oncology medicinal products in the European Union between 2017 and 2020. *Front Med (Lausanne)*. 2022;9:968272.
- Brettschneider C, Lühmann D, Raspe H. Informative value of patient reported outcomes (PRO) in health technology assessment (HTA). *GMS Health Technol Assess*. 2011;7:Doc01.
- Calvert M, Kyte D, Mercieca-Bebber R, et al. Guidelines for inclusion of patient-reported outcomes in clinical trial protocols: the SPIRIT-PRO extension. *JAMA*. 2018;319(5):483–494.
- Weldring T, Smith SM. Patient-reported outcomes (PROs) and patient-reported outcome measures (PROMs). *Health Serv Insights*. 2013;6:61–68.
- Rabin R, Gudex C, Selai C, Herdman M. From translation to version management: a history and review of methods for the cultural adaptation of the EuroQol five-dimensional questionnaire. *Value Health*. 2014;17(1):70–76.
- The EuroQol Group. EuroQol - a new facility for the measurement of health-related quality of life. *Health Policy*. 1990;16(3):199–208.
- Brooks R, the EuroQol Group. EuroQol: the current state of play. *Health Policy*. 1996;37(1):53–72.
- The EuroQol Group. EQ-5D. <https://euroqol.org/>. Accessed March 12, 2025.
- Herdman M, Gudex C, Lloyd A, et al. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Qual Life Res*. 2011;20(10):1727–1736.
- EQ-5D. EQ-5D-Y-5L EuroQol research foundation.
- Dolan PD. Modeling valuations for EuroQol health states. *Med Care*. 1997;35(11):1095–1108.
- Szende A, Oppe M, Devlin N. *EQ-5D Value Sets: Inventory, Comparative Review and User Guide*. Dordrecht, Netherlands: Springer; 2006.
- Shaw JW, Johnson JA, Coons SJ. US valuation of the EQ-5D health states development and testing of the D1 valuation model. *Med Care*. 2005;43(3):203–220.
- Stolk E, Ludwig K, Rand K, Van Hout B, Ramos-Goni J. Overview, update, and lessons learned from the International EQ-5D-5L valuation work: version 2 of the EQ-5D-5L valuation protocol. *Value Health*. 2019;22(1):23–30.
- Ramos-Goni J, Oppe M, Slaap B, Busschbach J, Stolk E. Quality control process for EQ-5D-5L valuation studies. *Value Health*. 2016;20(3):466–473.
- van Hout J, Janssen MF, Feng Y-S, et al. Interim scoring for the EQ-5D-5L: mapping the EQ-5D-5L to EQ-5D-3L value sets. *Value Health*. 2012;15(5):708–715.
- Devlin N, Parkin D, Janssen B. *Methods for Analysing and Reporting EQ-5D Data*. Cham, Switzerland: Springer; 2020.
- Rabin R, de Charro F. EQ-5D: a measure of health status from the EuroQol Group. *Ann Med*. 2001;33(5):337–343.
- Brazier JE, Rowen D, Lloyd A, Karimi M. Future directions in valuing benefits for estimating QALYs: is time up for the EQ-5D? *Value Health*. 2019;22(1):62–68.
- Wang A, Rand K, Yang Z, Brooks R, Busschbach J. The remarkably frequent use of EQ-5D in non-economic research. *Eur J Health Econ*. 2022;23(6):1007–1014.
- Szende A, Janssen B, Cabases J. *Self-Reported Population Health: An International Perspective based on EQ-5D*. Dordrecht, Netherlands: Springer; 2014.
- Covidence. Covidence: systematic review software. Veritas Health Innovation. <https://www.covidence.org/>. Accessed March 12, 2025.
- Gardiner JC, Luo Z, Roman LA. Fixed effects, random effects and GEE: what are the differences? *Stat Med*. 2009;28(2):221–239.
- Diggle P, Heagerty P, Liang K, Zeger S. *Analysis of Longitudinal Data*. 2nd ed. New York, NY: Oxford University Press; 2002.
- Hardin J, Hilbe J. *Generalized Estimating Equations*. Boca Raton, FL: Chapman & Hall. CRC; 2002.
- R: a language and environment for statistical computing. R Foundation for Statistical Computing. <https://www.R-project.org/>; Published 2024. Accessed March 12, 2025.
- Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*. 2021;372:n71.
- Cochrane Risk of Bias Working Group, RoB 2. A revised Cochrane risk-of-bias tool for randomized trials. <https://methods.cochrane.org/bias/resources/rob-2-revised-cochrane-risk-bias-tool-randomized-trials>. Accessed March 12, 2025.
- GRADE Working Group. GRADE handbook for grading quality of evidence and strength of recommendations. <https://gdt.gradepro.org/app/handbook/handbook.html>. Accessed March 12, 2025.
- Shaw C, Longworth L, Bennett B, McEntee-Richardson L, Shaw JW. A review of the use of EQ-5D for clinical outcome assessment in health technology assessment, regulatory claims, and published literature. *Patient*. 2023.
- Cheng LJ, Chen LA, Cheng JY, Herdman M, Luo N. Systematic review reveals that EQ-5D minimally important differences vary with treatment type and may decrease with increasing baseline score. *J Clin Epidemiol*. 2024;174:111487.
- Sayah A. A Systematic Literature Review of Important and Meaningful Differences in the EQ-5D index and visual analogue scale scores. *Value Health*. in press.
- Pickard AS, Neary MP, Cella D. Estimation of minimally important differences in EQ-5D utility and VAS scores in cancer. *Health Qual Life Outcomes*. 2007;5:70.
- Kim SK, Kim SH, Jo MW, Lee SI. Estimation of minimally important differences in the EQ-5D and SF-6D indices and their utility in stroke. *Health Qual Life Outcomes*. 2015;13:32.
- Marra CA, Woolcott JC, Kopec JA, et al. A comparison of generic, indirect utility measures (the HUI2, HUI3, SF-6D, and the EQ-5D) and disease-specific instruments (the RAQoL and the HAQ) in rheumatoid arthritis. *Soc Sci Med*. 2005;60(7):1571–1582.
- Tsiplova K, Pullenayegum E, Cooke T, Xie F. EQ-5D-derived health utilities and minimally important differences for chronic health conditions: 2011 Commonwealth Fund Survey of Sicker Adults in Canada. *Qual Life Res*. 2016;25(12):3009–3016.
- Hagiwara Y, Ohashi Y, Uesaka K, et al. Health-related quality of life of adjuvant chemotherapy with S-1 versus gemcitabine for resected pancreatic cancer: results from a randomised phase III trial (JASPAC 01). *Eur J Cancer*. 2018;93:79–88.
- Nauck MA, Buse JB, Mann JFE, et al. Health-related quality of life in people with type 2 diabetes participating in the LEADER trial. *Diabetes Obes Metab*. 2019;21(3):525–532.
- Humphries B, Cox JL, Parkash R, et al. Patient-reported outcomes and patient-reported experience of patients with atrial fibrillation in the Impact-AF clinical trial. *J Am Heart Assoc*. 2021;10(15):e019783.
- Pavesi M, Devlin N, Hakimi Z, et al. Understanding the effects on HR-QoL of treatment for overactive bladder: a detailed analysis of EQ-5D clinical trial data for mirabegron. *J Med Econ*. 2013;16(7):866–876.
- Dufvenberg M, Diarbakerli E, Charalampidis A, et al. Six-month results on treatment adherence, physical activity, spinal appearance, spinal deformity, and quality of life in an ongoing randomised trial on conservative treatment for adolescent idiopathic scoliosis (CONTRAS). *J Clin Med*. 2021;10(21):4967.
- Cheng LJ, Pan T, Chen LA, et al. The ceiling effects of EQ-5D-3L and 5L in general population health surveys: a systematic review and meta-analysis. *Value Health*. 2024.
- Konnopk A, Koenig H-H. The “no problems”-Problem: an empirical analysis of ceiling effects on the EQ-5D 5L. *Qual Life Res*. 2017;26(8):2079–2084.
- McBee M. Modeling outcomes with floor or ceiling effects: an introduction to the Tobit Model. *Gifted Child Q*. 2010;54(4):314–320.
- Twisk J, Rijnmen F. Longitudinal Tobit regression: a new approach to analyze outcome variables with floor or ceiling effects. *J Clin Epidemiol*. 2009;62(9):953–958.
- Austin PC, Escobar M, Kopec JA. The use of the Tobit Model for analyzing measures of health status. *Qual Life Res*. 2000;9(8):901–910.
- Sayers A, Whitehouse MR, Judge A, MacGregor AJ, Blom AW, Ben-Shlomo Y. Analysis of change in patient-reported outcome measures with floor and ceiling effects using the multilevel Tobit model: a simulation study and an example from a National Joint Register using body mass index and the Oxford Hip Score. *BMJ Open*. 2020;10(8):e033646.
- Grootendorst P. Censoring in statistical models of health status: what happens when one can do better than “1”. *Qual Life Res*. 2000;9(8):911–914.
- Di Tanna GL, Porter JK, Lipton RB, Hatswell AJ, Sapra S, Villa G. Longitudinal assessment of utilities in patients with migraine: an analysis of erenumab randomized controlled trials. *Health Qual Life Outcomes*. 2019;17(1):171.

53. Griffiths A, Paracha N, Davies A, Branscombe N, Cowie MR, Sculpher M. Analyzing health-related quality of life data to estimate parameters for cost-effectiveness models: an example using longitudinal EQ-5D data from the SHIFT randomized controlled trial. *Adv Ther.* 2017;34(3):753–764.
54. Pullenayegum EM, Wong HS, Childs A. Generalized additive models for the analysis of EQ-5D utility data. *Med Decis Mak.* 2013;33(2):244–251.
55. Hernandez 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 Health.* 2012;15(3):550–561.
56. Alava MH, Wailoo A. Fitting adjusted limited dependent variable mixture models to EQ-5D. *STATA J.* 2015;15(3):737–750.
57. Vickers AJ, Altman DG. Statistics Notes: analysing controlled trials with baseline and follow up measurements. *BMJ.* 2001;323:7321.
58. Senn S, Julious S. Measurement in clinical trials: a neglected issue for statisticians? *Stat Med.* 2009;28(26):3189–3209.
59. Hernán MA, Hernández-Díaz S. Beyond the intention to treat in comparative effectiveness research. *Clin Trials.* 2012;9(1):48–55.
60. Vickers AJ. The use of percentage change from baseline as an outcome in a controlled trial is statistically inefficient: a simulation study. *BMC Med Res Methodol.* 2001;1:6.
61. Clifton L, Clifton DA. The correlation between baseline score and post-intervention score, and its implications for statistical analysis. *Trials.* 2019;20(1):43.
62. Adjusting for covariates in randomized clinical trials for drugs and biological products. Department of Health and Human Services, US Food and Drug Administration. <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/adjusting-covariates-randomized-clinical-trials-drugs-and-biological-products>; Published 2023. Accessed March 12, 2025.
63. Guideline on adjustment for baseline covariates in clinical trials. European Medicines Agency. https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-adjustment-baseline-covariates-clinical-trials_en.pdf; Published 2015. Accessed March 12, 2025.
64. Coens C, Pe M, Dueck AC, et al. International standards for the analysis of quality-of-life and patient-reported outcome endpoints in cancer randomised controlled trials: recommendations of the SISAQOL Consortium. *Lancet Oncol.* 2020;21(2):e83–e96.
65. National Research Council (US) Panel on Handling Missing Data in Clinical Trials. *The Prevention and Treatment of Missing Data in Clinical Trials.* Washington, DC: National Academies Press; 2011.
66. Little RJA, Rubin DB. *Statistical Analysis With Missing Data.* 3rd ed. Hoboken, NJ: Wiley, John Wiley & Sons, Inc; 2020.
67. Mallinckrodt CH, Lane PW, Schnell D, Peng Y, Mancuso JP. Recommendations for the primary analysis of continuous endpoints in longitudinal clinical trials. *Statistics.* 2008;42:303–319.
68. Pe M, Dorme L, Coens C, et al. Statistical analysis of patient-reported outcome data in randomised controlled trials of locally advanced and metastatic breast cancer: a systematic review. *Lancet Oncol.* 2018;19(9):e459–e469.
69. Qian Y, Walters SJ, Jacques R, Flight L. Comprehensive review of statistical methods for analysing patient-reported outcomes (PROs) used as primary outcomes in randomised controlled trials (RCTs) published by the UK's health technology assessment (HTA). *BMJ Open.* 2021;11(9):e051673.
70. Cnaan A, Laird NM, Slasor P. Using the general linear mixed model to analyse unbalanced repeated measures and longitudinal data. *Stat Med.* 1997;16(20):2349–2380.
71. Scharfstein DO, McDermott A. Global sensitivity analysis of clinical trials with missing patient-reported outcomes. *Stat Methods Med Res.* 2019;28(5):1439–1456.
72. Holzhauer B, Akacha M, Bermann G. Choice of estimand and analysis methods in diabetes trials with rescue medication. *Pharm Stat.* 2015;14(6):433–447.
73. Little RJ, Lewis RJ. Estimators, and estimates. *JAMA.* 2021;326(10):967–968.
74. Kahan BC, Morris TP, White IR, Carpenter J, Cro S. Estimands in published protocols of randomised trials: urgent improvement needed. *Trials.* 2021;22(1):686.
75. Mallinckrodt C, Molenberghs G, Rathmann S. Choosing estimands in clinical trials with missing data. *Pharm Stat.* 2017;16(1):29–36.
76. Pohl M, Baumann L, Behnisch R, Kirchner M, Krisam J, Sander A. Estimands—a basic element for clinical trials. *Dtsch Arztebl Int.* 2021;118(51–52):883–888.
77. Kahan BC, Hindley J, Edwards M, Cro S, Morris TP. The estimands framework: a primer on the ICH E9(R1) addendum. *BMJ.* 2024;384:e076316.
78. Gogtay NJ, Ranganathan P, Aggarwal R. Understanding estimands. *Perspect Clin Res.* 2021;12(2):106–112.