Study Protocol and Statistical Analysis Plan: Construct validity and responsiveness of EQ-5D-3L and EQ-5D-5L in patients with inflammatory joint disease

2026-04-20

Karolinska Institutet

Karolinska Institutet

Table of contents

Brief summary

The aim of this study is to compare the construct validity (convergent validity and known-groups validity) and responsiveness of EQ-5D-3L and EQ-5D-5L in patients with inflammatory joint disease. The study is based on prospectively collected data through the Swedish Rheumatology Quality Register (SRQ).

Background

EQ-5D is a generic instrument used to measure, value, and compare health and health-related quality of life (HRQoL) across symptoms and diagnoses [1,2]. There are two versions of the EQ-5D, one with three response levels to each item (EQ-5D-3L) and one with five response levels to each item (EQ-5D-5L) [1–3]. EQ-5D-5L was developed in response to concerns about the responsiveness of the EQ-5D-3L [3].

When choosing an instrument to follow-up care, it is important to know that the instrument has good construct validity and responsiveness [4,5]. An instrument with low construct validity or responsiveness might not capture and describe patients' self-perceived health state and changes in self-perceived health state accurately. Previous research done by the authors on patients with inflammatory joint disease has shown that the EQ-5D-3L has low responsiveness [6]. However, the comparative responsiveness and construct validity of the EQ-5D-3L and EQ-5D-5L in patients with inflammatory joint disease are unknown. The aim of this study is therefore to assess and compare the construct validity and responsiveness of the EQ-5D-3L and the EQ-5D-5L in patients with inflammatory joint disease. The research questions are:

1. What is the construct validity of EQ-5D-3L and EQ-5D-5L in patients with inflammatory joint disease?
2. What is the responsiveness of EQ-5D-3L and EQ-5D-5L in patients with inflammatory joint disease?
3. How does the performance of EQ-5D-5L compare with EQ-5D-3L in terms of construct validity and responsiveness in patients with inflammatory joint disease?

Methodology

Study design

The study will be based on prospectively collected data. The study will follow the established guidelines from COSMIN on how to assess construct validity and responsiveness of patient-reported outcome measures (PROMs) [5,7–9]. Construct validity refers to the degree to which an instrument measures the constructs it intends to measure [5].

Construct validity will be assessed in two ways: convergent validity and known-groups validity. Convergent validity refers to how well an instrument (here EQ-5D) correlates with comparator instruments measuring similar or related constructs [5]. Preferably, the comparator instrument would be a gold standard; however, when measuring abstract constructs such as health, gold standards rarely exist. Therefore, measures of similar and related constructs are used. According to the COSMIN guidelines, convergent validity should be assessed by formulating and testing hypotheses regarding the expected direction and magnitude of the correlation between the instrument being studied and the comparator instruments measuring similar or related constructs [5,8,9]. Known-groups validity refers to how well the instrument can identify differences between groups known to differ [5,9]. Known-groups validity should be assessed by formulating and testing hypotheses regarding expected directions and magnitude of the differences between subgroups [9].

Responsiveness refers to the ability of an instrument (here EQ-5D) to capture change over time in the construct that is measured [5]. According to COSMIN, responsiveness can be assessed by comparing changes in scores generated using the instrument under study with changes in scores generated using the comparator instruments, similar to the principles explained above for convergent validity [5,8,9]. Responsiveness can also be assessed by analysing whether the instrument under study can discriminate between patients who have improved and those who have not, based on changes in the comparator instrument, such as changes in disease activity or physical function [5,9].

Ethical approval has been granted for the project (2023-04394-01, and ethical amendments 2023-04362-02 and 2025-06429-02).

Data collection

Data on EQ-5D-5L is prospectively collected through an online platform called Patients Own Registration (PER). Data collection started in November 2023 and concluded in March 2026. The study includes 6,967 patients with inflammatory joint disease. Patients with inflammatory joint disease already complete PROMs, including EQ-5D-3L, routinely in conjunction with their follow-up healthcare visits, and patients who consent to participate in the study will also complete the EQ-5D-5L. The order in which the patient responds to the EQ-5D-3L and EQ-5D-5L is randomised so that participants answer one version of the EQ-5D before the other routinely collected PROMs, and the other EQ-5D version at the end. When data collection with EQ-5D-5L is concluded, all data necessary for analyses will be retrieved from the SRQ.

Study population

Inclusion criteria are:

- Agreed to participate in the study
- ≥18 years at the time of the first measurement
- A diagnosis of rheumatoid arthritis (RA), polyarthritis, psoriatic arthritis (PsA), or ankylosing spondylitis (AS)
- Complete registration of responses in the EQ-5D-3L and EQ-5D-5L descriptive system at least at one time point (one visit)
- For patients with RA: At least one measurement with Disease Activity Score 28 (DAS28) reported in relation to the same visit as the EQ-5D was registered
- For patients with polyarthritis: At least one measurement with DAS28 reported in relation to the same visit as the EQ-5D was registered
- For patients with PsA: At least one measurement with DAS28 or Disease Activity in Psoriatic Arthritis (DAPSA) reported in relation to the same visit as the EQ-5D was registered
- For patients with AS: At least one measurement with the Bath Ankylosing Spondylitis Disease Activity Index (BASDAI) or the Axial Spondyloarthritis Disease Activity Score (ASDAS) reported in relation to the same visit as the EQ-5D was registered

For the analyses of construct validity, the first measurement will be used if the patient has multiple complete registrations with EQ-5D and the comparator instrument. The hypotheses for responsiveness will be tested in patients with newly diagnosed disease (having the diagnosis for ≤12

months), as changes in health are most likely to be present in this group. For the responsiveness analysis, we will use two visits from the first year: the first complete visit and the complete visit that occurs closest to after six months after the first visit, so that there is at least six months between visit one and visit two.

Outcome measures

EQ-5D-3L and EQ-5D-5L

The EQ-5D-3L and the EQ-5D-5L measure HRQoL and consist of two parts [1–3]. The first part, the descriptive system, contains five questions, each relating to a distinct HRQoL dimension: mobility, usual activities, self-care, pain/discomfort, and anxiety/depression. In the EQ-5D-3L version, each question can be answered with no problems (level 1), some/moderate problems (level 2), or unable to perform certain activities/having extreme problems (level 3). In the EQ-5D-5L version, each question can be answered with no problems (level 1), slight problems (level 2), moderate problems (level 3), severe problems (level 4), or extreme problems (level 5) [3]. There are also differences in the wording of some dimensions between the EQ-5D-3L and the EQ-5D-5L, such as confined to bed for level 3 mobility (EQ-5D-3L) versus unable to walk about for level 5 (EQ-5D-5L). The responses to the descriptive system can be summarised into an index value based on an existing preference-based value set. In this study, the EQ-5D-3L value set by Dolan [10] will be used. For the EQ-5D-5L, the EQ-5D-5L value set by Rowen et al. [11] will be used for the main analyses. Other value sets, such as the Swedish EQ-5D-5L value set by Sun et al. [12] and the UK crosswalk from EQ-5D-3L to EQ-5D-5L by Hernández Alava et al. [13] or van Hout et al. [14], can be used for sensitivity analyses.

The second part of the EQ-5D is EQ VAS, a visual analogue scale, where the respondent records their self-rated health today on a scale from 0 (worst imaginable health) to 100 (best imaginable health) [1,2].

Comparator instruments for disease activity

DAS28 was developed for patients with RA and contains questions about tender and swollen joints, disease activity, and inflammation [15]. DAPSA was developed for patients with PsA and contains questions about tender and swollen joints, disease activity, pain, and inflammation [16,17]. BASDAI was

developed to measure disease activity in patients with AS and contains questions about fatigue, pain, swelling, and stiffness [18,19]. ASDAS was developed to measure disease activity in patients with AS and contains questions about pain, swelling, stiffness, disease activity, and inflammation [19].

Common to all instruments, a higher score indicates higher disease activity.

Comparator instruments for physical function
The Health Assessment Questionnaire Disability Index (HAQ-DI) [20] is a generic instrument and the Bath Ankylosing Spondylitis Functional Index (BASFI) [21] is a disease-specific instrument for use in AS. Both measure physical function through questions about mobility, daily activities, and self-care. For both instruments, a higher value indicates more problems with physical function.

Comparator instruments for pain, fatigue, and general health
The SRQ also contains three questions about pain, fatigue, and general health, measured with a visual analogue scale (VAS) [22]. The items assess how much pain the person has had, how fatigued they have been, and how their general health has been in the last week due to their rheumatic disease. The scales are horizontal and range from no problem (0) to as bad as it can be (100) [22].

Analyses of construct validity and responsiveness
To assess the construct validity and responsiveness, a priori hypotheses for convergent validity, known-groups validity, and responsiveness have been formulated. To support construct validity or responsiveness within each patient group, at least 75% of the hypotheses specified below must be supported by the data [5,7]. To determine whether EQ-5D-3L or EQ-5D-5L demonstrates superior construct validity or responsiveness, the proportions of supported hypotheses will be compared. The instrument with the higher proportion of supported hypotheses will be considered to exhibit better construct validity or responsiveness.

Hypotheses relating to the convergent validity of EQ-5D
Convergent validity will be assessed by testing hypotheses regarding the expected direction and magnitude of correlation between EQ-5D scores

and scores generated using the comparator instruments. Constructs that are considered to be related are expected to have at least a moderate correlation of ≥0.3, and constructs that are considered to be similar are expected to have a strong correlation of≥0.5 [8]. The hypotheses for all patient groups are presented in the tables (1-5) below.

Table 1. Hypotheses for convergent validity analyses based on VAS pain

Instruments or items being analysed		Theoretical relationship	Expected correlation
VAS pain	EQ-5D index	Related	At least a moderate negative correlation
VAS pain	EQ-5D **mobility**	Related	At least a moderate positive correlation
VAS pain	EQ-5D **self-care**	Related	At least a moderate positive correlation
VAS pain	EQ-5D **usual activities**	Related	At least a moderate positive correlation
VAS pain	EQ-5D **pain/discomfort**	Similar	At least a strong positive correlation

Table 2. Hypotheses for convergent validity analyses based on VAS general health

Instruments or items being analysed		Theoretical relationship	Expected correlation
VAS general health	EQ-5D index	Similar	At least a strong negative correlation

Table 3. Hypotheses for convergent validity analyses based on VAS fatigue

Instruments or items being analysed		Theoretical relationship	Expected correlation
VAS fatigue	EQ-5D index	Related	At least a moderate negative correlation
VAS fatigue	EQ-5D **mobility**	Related	At least a moderate positive correlation
VAS fatigue	EQ-5D **self-care**	Related	At least a moderate positive correlation

VAS fatigue	EQ-5D **usual activities**	Related	At least a moderate positive correlation
VAS fatigue	EQ-5D **anxiety/depression**	Related	At least a moderate positive correlation

Table 4. Hypotheses for convergent validity analyses based on disease activity measured with (DAS28/DAPSA/BASDAI/ASDAS)

Instruments or items being analysed		Theoretical relationship	Expected correlation
Disease activity (DAS28/DAPSA/ BASDAI/ASDAS)	EQ-5D index	Related	At least a moderate negative correlation
Disease activity (DAS28/DAPSA/ BASDAI/ASDAS)	EQ-5D **mobility**	Related	At least a moderate positive correlation
Disease activity (DAS28/DAPSA/ BASDAI/ASDAS)	EQ-5D **self-care**	Related	At least a moderate positive correlation
Disease activity (DAS28/DAPSA/ BASDAI/ASDAS)	EQ-5D **usual activities**	Related	At least a moderate positive correlation
Disease activity (DAS28/DAPSA/ BASDAI/ASDAS)	EQ-5D **pain/ discomfort**	Related	At least a moderate positive correlation

Table 5. Hypotheses for convergent validity analyses based on physical function measured with (HAQ-DI/BASFI)

Instruments or items being analysed		Theoretical relationship	Expected correlation
Physical function (HAQ-DI/BASFI)	EQ-5D index	Related	At least a moderate negative correlation
Physical function (HAQ-DI/BASFI)	EQ-5D **mobility**	Related	At least a moderate positive correlation
Physical function (HAQ-DI/BASFI)	EQ-5D **self-care**	Related	At least a moderate positive correlation
Physical function (HAQ-DI/BASFI)	EQ-5D **usual activities**	Related	At least a moderate positive correlation
Physical function (HAQ-DI/BASFI)	EQ-5D **pain/ discomfort**	Related	At least a moderate positive correlation

Hypotheses relating to the known-groups validity of EQ-5D

In the assessment of known-groups validity, patients will be divided into groups for which there is an expected difference in HRQoL. The groups represent patients with different levels of disease activity or physical function. Patients with lower disease activity are expected to have better HRQoL compared to patients with higher disease activity. Likewise, patients with better physical function are expected to have better HRQoL compared to patients with worse physical function. The hypotheses are:

- Patients with low disease activity will have higher EQ-5D index values
- Patients with low disease activity will report fewer problems in the dimensions for mobility, self-care, usual activities, and pain/discomfort.
- Patients with high physical function will have higher EQ-5D index values
- Patients with high physical function will report fewer problems in the dimensions for mobility, self-care, usual activities, and pain/discomfort.

The cut-off values for disease activity (DAS28 [23], DAPSA [24], ASDAS [24], BASDAI [24–26]) and physical function (HAQ-DI [27,28], BASFI [24,25]) are based on prior literature (Table 6). The literature uses different terminology when classifying patients according to how active their disease is. For the purpose of this study protocol, the groups are called low disease activity and high disease activity. For DAS28, the group with low disease activity will include patients with remission and low disease activity and the group with high disease activity will include patients with moderate and high disease activity [23]. For DAPSA, the group with low disease activity will include patients in remission and with low disease activity and the group with high disease activity will include patients with moderate and high disease activity. For ASDAS, the group with low disease activity will include patients with inactive and low disease activity and the group with high disease activity will include patients with high and very high disease activity. BASDAI ≥4 has been mentioned in the literature as active disease and this cut-off value has been used in another study for known-groups validity [24–26].

For HAQ-DI, the group with high physical function will include patients with mild to moderate difficulty and the group with low physical function will include patients with moderate to severe disability [27,28]. BASFI will be divided based on the same cut-off values as BASDAI.

Table 6. Cut-off values used to group patients for known-groups validity

Measure	**Low** disease activity	**High** disease activity
Disease activity DAS28	<3.2	≥3.2
Disease activity DAPSA	≤14	>14
Disease activity ASDAS	<2.1	≥2.1
Disease activity BASDAI	<4	≥4
	High physical function	**Low** physical function
Physical function HAQ-DI	<1	≥1
Physical function BASFI	<4	≥4

To estimate whether there is a difference in the EQ-5D mean index value between the groups, the effect size (ES) will be calculated. Cohen's D will be calculated and the ES is considered small if it is ≥0.2, medium if it is ≥0.5, and large if it is ≥0.8 [29]. To support the hypotheses for the EQ-5D index, at least a moderate ES will be required.

To examine differences in the distribution of individual-level ordinal responses within the EQ-5D descriptive system, the Mann–Whitney U test will be applied. This non-parametric approach is chosen to account for the ordinal data. Moreover, the *r* ES will be calculated. The *r* ES is considered small if it is ≥0.1, medium if it is ≥0.3, and large if it is ≥0.5 [29,30]. To support the hypotheses for the dimensions, at least a moderate ES will be required.

Hypotheses relating to the responsiveness of EQ-5D

Responsiveness will be assessed in two ways. Firstly, by assessing the relationship between individual changes in EQ-5D index value and dimensions over time, with changes in the comparator instruments over the same time period. The second way is to assess whether the EQ-5D index can discriminate between patients who have improved over time and those who have not, based on changes in disease activity or physical function.

Correlation between changes in EQ-5D scores and changes in scores generated using comparator instruments

The relationship between changes in EQ-5D scores and changes in scores generated using comparator instruments will be assessed by analysing the correlation between score changes between two visits during the first year of the diagnosis. The hypotheses for expected correlations are presented in Tables 7-11 below.

Table 7. Hypotheses for analyses of responsiveness based on VAS pain

Instruments or items being analysed		Theoretical relationship	Expected correlation
VAS pain	EQ-5D index	Related	At least a moderate negative correlation
VAS pain	EQ-5D **mobility**	Related	At least a moderate positive correlation
VAS pain	EQ-5D **self-care**	Related	At least a moderate positive correlation
VAS pain	EQ-5D **usual activities**	Related	At least a moderate positive correlation
VAS pain	EQ-5D **pain/discomfort**	Similar	At least a strong positive correlation

Table 8. Hypotheses for analyses of responsiveness based on VAS general health

Instruments or items being analysed		Theoretical relationship	Expected correlation
VAS general health	EQ-5D index	Similar	At least a strong negative correlation

Table 9. Hypotheses for analyses of responsiveness based on VAS fatigue

Instruments or items being analysed		Theoretical relationship	Expected correlation
VAS fatigue	EQ-5D index	Related	At least a moderate negative correlation
VAS fatigue	EQ-5D **mobility**	Related	At least a moderate positive correlation
VAS fatigue	EQ-5D **self-care**	Related	At least a moderate positive correlation
VAS fatigue	EQ-5D **usual activities**	Related	At least a moderate positive correlation

VAS fatigue	EQ-5D **anxiety/depression**	Related	At least a moderate positive correlation

Table 10. Hypotheses for analyses of responsiveness based on disease activity measured with DAS28/DAPSA/BASDAI/ASDAS

Instruments or items being analysed		Theoretical relationship	Expected correlation
Disease activity (DAS28/DAPSA/ BASDAI/ASDAS)	EQ-5D index	Related	At least a moderate negative correlation
Disease activity (DAS28/DAPSA/ BASDAI/ASDAS)	EQ-5D **mobility**	Related	At least a moderate positive correlation
Disease activity (DAS28/DAPSA/ BASDAI/ASDAS)	EQ-5D **self-care**	Related	At least a moderate positive correlation
Disease activity (DAS28/DAPSA/ BASDAI/ASDAS)	EQ-5D **usual activities**	Related	At least a moderate positive correlation
Disease activity (DAS28/DAPSA/ BASDAI/ASDAS)	EQ-5D **pain/ discomfort**	Related	At least a moderate positive correlation

Table 11. Hypotheses for analyses of responsiveness based on physical function measured with HAQ-DI/BASFI

Instruments or items being analysed		Theoretical relationship	Expected correlation
Physical function (HAQ-DI/BASFI)	EQ-5D index	Related	At least a moderate negative correlation
Physical function (HAQ-DI/BASFI)	EQ-5D **mobility**	Related	At least a moderate positive correlation
Physical function (HAQ-DI/BASFI)	EQ-5D **self-care**	Related	At least a moderate positive correlation
Physical function (HAQ-DI/BASFI)	EQ-5D **usual activities**	Related	At least a moderate positive correlation
Physical function (HAQ-DI/BASFI)	EQ-5D **pain/ discomfort**	Related	At least a moderate positive correlation

Analysing EQ-5D in patients who have improved versus patients who have not improved based on disease activity or physical function

To assess whether the EQ-5D index value can discriminate between patients who have improved and those who have not, the area under the receiver operating characteristic curve (AUC) will be calculated. Patients will be considered to have improved based on results from the comparator instruments using criteria from the literature for improvement definitions (Table 12). To determine whether EQ-5D can discriminate between patients who have improved and those who have not, the AUC is required to be ≥0.70 [5,8,9]. The hypotheses for these analyses are presented below.

DAS28/DAPSA/ASDAS/BASDAI

- The EQ-5D **index** is expected to be able to discriminate between patients who have improved disease activity and patients who have not improved.

BASFI/HAQ-DI

- The EQ-5D **index** is expected to be able to discriminate between patients who have improved physical function and patients who have not improved.

Table 12. Improvement criteria

Instruments	Improvement criteria	Reference
BASDAI and BASFI	>50% or more than two points improvement on the scale	[24]
ASDAS CRP	≥1.1 points improvement on the scale	[24]
DAPSA	≥50% improvement	[31]
DAS28	>1.2 if DAS28 is >5.1 at the endpoint and >0.6 if DAS28 is ≤5.1 at the endpoint	[32]
HAQ-DI	≥0.22 points improvement on the scale	[33]

Hypotheses relating to the convergent validity of EQ VAS

Convergent validity of EQ VAS will be assessed with the same analyses used for EQ-5D. The hypotheses for these analyses are presented below (Table 13).

Table 13. Hypotheses for analyses of convergent validity of EQ VAS

Instruments or items being analysed		Theoretical relationship	Expected correlation
VAS general health	EQ VAS	**Similar**	At least a strong negative correlation
VAS pain	EQ VAS	**Related**	At least a moderate negative correlation
VAS fatigue	EQ VAS	**Related**	At least a moderate negative correlation
Disease activity (DAS28/DAPSA/ BASDAI/ASDAS)	EQ VAS	**Related**	At least a moderate negative correlation
Physical function (HAQ-DI/BASFI)	EQ VAS	**Related**	At least a moderate negative correlation

Hypotheses relating to the known-groups validity of EQ VAS
Known-groups validity of EQ VAS will be assessed for the same groups as the EQ-5D index (see section about known-groups validity EQ-5D). To estimate whether there is a difference in EQ VAS between the groups, Cohen's D will be calculated. To support the hypotheses for the EQ VAS, at least a moderate ES will be required. The hypotheses are:

- Patients with low disease activity will have higher EQ VAS
- Patients with high physical function will have higher EQ VAS

Hypotheses relating to the responsiveness of EQ VAS
Responsiveness for EQ VAS will be assessed with the same analyses used for EQ-5D (see the section about responsiveness EQ-5D). The hypotheses are described below.

Hypotheses for correlations between changes in EQ VAS scores and changes in scores generated using comparator instruments
The relationship between changes in EQ VAS scores and changes in scores generated using comparator instruments will be assessed by analysing the correlation between score changes between two visits during the first year of diagnosis. The hypotheses for expected correlations are presented in Table 14.

Table 14. Hypotheses for analyses of responsiveness of EQ VAS

Instruments or items being analysed		Theoretical relationship	Expected correlation
VAS general health	EQ VAS	**Similar**	At least a strong negative correlation
VAS pain	EQ VAS	**Related**	At least a moderate negative correlation
VAS fatigue	EQ VAS	**Related**	At least a moderate negative correlation
Disease activity (DAS28/DAPSA/ BASDAI/ASDAS)	EQ VAS	**Related**	At least a moderate negative correlation
Physical function (HAQ-DI/BASFI)	EQ VAS	**Related**	At least a moderate negative correlation

Hypotheses for analysing EQ VAS in patients who have improved versus patients who have not improved based on disease activity or physical function

To determine whether EQ VAS can discriminate between patient groups, the AUC is required to be ≥0.70 [5,8,9]. The hypotheses for these analyses are presented below.

DAS28/DAPSA/ASDAS/BASDAI

- EQ VAS is expected to be able to discriminate between patients who have **improved** disease activity and patients who have not improved.

BASFI/HAQ-DI

- EQ VAS is expected to be able to discriminate between patients who have **improved** physical function and patients who have not improved.

Descriptive Analyses

In addition to the analyses of construct validity and responsiveness of the two EQ-5D versions based on hypotheses, descriptive analyses will be conducted. These are commonly applied analyses in EQ-5D research to evaluate psychometric and distributional properties (e.g., clustering, gaps,

skewness, ceiling effects) but are not included in the guidelines by COSMIN. Thus, they will not influence our conclusions regarding whether the findings of the study support the construct validity and responsiveness of EQ-5D-3L and EQ-5D-5L. However, they will be used to interpret and discuss the findings from the hypothesis-driven analyses.

Cross-sectionally

Distributional characteristics of the two versions of EQ-5D will be studied using quantitative and graphical methods. Quantitative analyses will compare EQ-5D-3L and EQ-5D-5L with respect to ceiling and floor effects, as well as the distribution of health profiles. Ceiling and floor will be defined as the proportion of respondents reporting the best and the worst health state, respectively, and will be assessed at the dimension level and for index values. The proportion of observed health profiles will be calculated as the number of unique profiles observed divided by the total number of possible profiles. Graphical analysis will examine the distribution of index values with particular attention to clusters and gaps.

Moreover, informativity will be analysed using Shannon's indices, with the Shannon index (H') for absolute and Shannon Evenness index (J') for relative informativity [34]. The Shannon index (H') reflects the overall spread of responses across response levels and increases with both the number of levels used and how evenly they are used. In contrast, the Shannon Evenness index (J') expresses how evenly responses are distributed relative to the maximum possible spread, independent of the number of response categories. Higher values of both indices indicate a greater ability of the instrument to distinguish between health states. Informativity will be examined both at the dimension level and for the index values. Because J' standardises H' to its maximum possible value, it facilitates comparison of informativity between instruments with different numbers of response levels, such as the EQ-5D-3L and EQ-5D-5L.

Inconsistency of responses will be assessed using criteria established in previous studies [35–37], where an inconsistent response is defined as an EQ-5D-3L response followed by an EQ-5D-5L response that differs by at least two levels. Inconsistency of responses will also be explored graphically

to illustrate response patterns across corresponding dimensions on the EQ-5D-3L and the EQ-5D-5L.

Longitudinal analyses

Several descriptive longitudinal analyses will be conducted. First, inconsistencies between EQ-5D-3L and EQ-5D-5L (EQ-5D-3L improves on a certain dimension while EQ-5D-5L worsens, and vice versa) will be assessed.

Second, changes in health states will be classified using the Paretian Classification of Health Change (PCHC), categorising patients as improved, worsened, mixed, or unchanged [38]. Improvement is defined as improvement in at least one dimension with no deterioration in any other dimension, while worsening is defined as deterioration in at least one dimension with no improvement elsewhere. Mixed change indicates simultaneous improvement in one or more dimensions and deterioration in one or more other dimensions. No change reflects identical responses across all dimensions.

Third, we will calculate the probability of superiority, a non-parametric effect size measure [39], per EQ-5D dimension. The probability of superiority will be calculated as the number of patients who improve, plus half the number of patients who remained unchanged, divided by the total number of patients. Values below 0.5 indicate more patients deteriorated than improved, whereas values above 0.5 indicate more improved than deteriorated.

Lastly, we will calculate the standardized response mean (SRM) and standardized effect size (SES) for the EQ-5D-3L and EQ-5D-5L index values using disease activity and physical function as external indicators for change. The relative efficiency of EQ-5D-5L over EQ-5D-3L will be assessed by calculating ratios of the SRM and SES statistics. We will also calculate confidence intervals using bootstrapping. The comparator instruments measuring disease activity and physical function will be used to classify patients into improved and not improved.

References

[1] Brooks R. EuroQol: the current state of play. Health Policy Amst Neth 1996;37:53–72. https://doi.org/10.1016/0168-8510(96)00822-6.

[2] EuroQol Group. EuroQol--a new facility for the measurement of health-related quality of life. Health Policy Amst Neth 1990;16:199–208. https://doi.org/10.1016/0168-8510(90)90421-9.

[3] Herdman M, Gudex C, Lloyd A, Janssen MF, Kind P, Parkin D, et al. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). Qual Life Res 2011;20:1727–36. https://doi.org/10.1007/s11136-011-9903-x.

[4] Streiner DL, Norman GR, Cairney J. Health Measurement Scales: A Practical Guide to Their Development and Use. Oxford, UNITED KINGDOM: Oxford University Press, Incorporated; 2015.

[5] Mokkink LB, Elsman EBM, Terwee CB. COSMIN guideline for systematic reviews of patient-reported outcome measures version 2.0. Qual Life Res Int J Qual Life Asp Treat Care Rehabil 2024;33:2929–39. https://doi.org/10.1007/s11136-024-03761-6.

[6] Degerlund-Maldi K, Regardt M, Nystrand Länsman C, Larsson L, Parodis I, Heintz E. The value of EQ-5D-3L and EQ VAS as a patient-reported outcome measure for patients with ankylosing spondylitis in routine healthcare: an evaluation of construct validity and responsiveness based on the Swedish Rheumatology Quality Register. J Patient-Rep Outcomes 2026. https://doi.org/10.1186/s41687-026-01009-0.

[7] Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol 2007;60:34–42. https://doi.org/10.1016/j.jclinepi.2006.03.012.

[8] Prinsen CAC, Mokkink LB, Bouter LM, Alonso J, Patrick DL, de Vet HCW, et al. COSMIN guideline for systematic reviews of patient-reported outcome measures. Qual Life Res Int J Qual Life Asp Treat Care Rehabil 2018;27:1147–57. https://doi.org/10.1007/s11136-018-1798-3.

[9] Mokkink LB, Prinsen CA, Patrick DL, Alonso J, Bouter LM, de Vet HC, et al. COSMIN Study Design checklist for Patient-reported outcome measurement instruments 2019.

[10] Dolan P. Modeling valuations for EuroQol health states. Med Care 1997;35:1095–108. https://doi.org/10.1097/00005650-199711000-00002.

[11] Rowen D, Mukuria C, Bray N, Carlton J, Longworth L, Meads D, et al. A UK value set for the EQ-5D-5L. Value Health 2026. https://doi.org/10.1016/j.jval.2026.03.008.

[12] Sun S, Chuang L-H, Sahlén K-G, Lindholm L, Norström F. Estimating a social value set for EQ-5D-5L in Sweden. Health Qual Life Outcomes 2022;20:167. https://doi.org/10.1186/s12955-022-02083-w.
[13] Hernández Alava M, Pudney S, Wailoo A. Estimating the Relationship Between EQ-5D-5L and EQ-5D-3L: Results from a UK Population Study. PharmacoEconomics 2023;41:199–207. https://doi.org/10.1007/s40273-022-01218-7.
[14] van Hout B, Janssen MF, Feng Y-S, Kohlmann T, Busschbach J, Golicki D, et al. Interim Scoring for the EQ-5D-5L: Mapping the EQ-5D-5L to EQ-5D-3L Value Sets. Value Health 2012;15:708–15. https://doi.org/10.1016/j.jval.2012.02.008.
[15] Prevoo ML, van 't Hof MA, Kuper HH, van Leeuwen MA, van de Putte LB, van Riel PL. Modified disease activity scores that include twenty-eight-joint counts. Development and validation in a prospective longitudinal study of patients with rheumatoid arthritis. Arthritis Rheum 1995;38:44–8. https://doi.org/10.1002/art.1780380107.
[16] Nell-Duxneuner VP, Stamm TA, Machold KP, Pflugbeil S, Aletaha D, Smolen JS. Evaluation of the appropriateness of composite disease activity measures for assessment of psoriatic arthritis. Ann Rheum Dis 2010;69:546–9. https://doi.org/10.1136/ard.2009.117945.
[17] Schoels M, Aletaha D, Funovits J, Kavanaugh A, Baker D, Smolen JS. Application of the DAREA/DAPSA score for assessment of disease activity in psoriatic arthritis. Ann Rheum Dis 2010;69:1441–7. https://doi.org/10.1136/ard.2009.122259.
[18] Calin A, Nakache J-P, Gueguen A, Zeidler H, Mielants H, Dougados M. Defining disease activity in ankylosing spondylitis: is a combination of variables (Bath Ankylosing Spondylitis Disease Activity Index) an appropriate instrument? Rheumatology 1999;38:878–82. https://doi.org/10.1093/rheumatology/38.9.878.
[19] Lukas C, Landewé R, Sieper J, Dougados M, Davis J, Braun J, et al. Development of an ASAS-endorsed disease activity score (ASDAS) in patients with ankylosing spondylitis. Ann Rheum Dis 2009;68:18. https://doi.org/10.1136/ard.2008.094870.
[20] Fries JF, Spitz P, Kraines RG, Holman HR. Measurement of patient outcome in arthritis. Arthritis Rheum 1980;23:137–45. https://doi.org/10.1002/art.1780230202.
[21] Calin A, Garrett S, Whitelock H, Kennedy LG, O'Hea J, Mallorie P, et al. A new approach to defining functional ability in ankylosing spondylitis: the development of the Bath Ankylosing Spondylitis Functional Index. J Rheumatol 1994;21:2281–5.

[22] Svensk Reumatologis Kvalitetsregister. BESÖKSBLANKETT - Patient. 2023. https://srq.nu/wp-content/uploads/2023/10/Besoksblankett-alla-diagnoser.pdf (accessed July 31, 2025).
[23] Aletaha D, Smolen JS. The Simplified Disease Activity Index (SDAI) and Clinical Disease Activity Index (CDAI) to monitor patients in standard clinical care. Best Pract Res Clin Rheumatol 2007;21:663–75. https://doi.org/10.1016/j.berh.2007.02.004.
[24] Svensk Reumatologisk Förening. Riktlinjer för läkemedelsbehandling vid axial spondylartrit och psoriasisartrit 2022 2022. https://svenskreumatologi.se/wp-content/uploads/2023/01/riktlinjer-axial-spondylartrit-och-psoriasisartrit-220127.pdf (accessed May 8, 2024).
[25] Boonen A, van der Heijde D, ewé R, van Tubergen A, Mielants H, Dougados M, et al. How do the EQ-5D, SF-6D and the well-being rating scale compare in patients with ankylosing spondylitis? Ann Rheum Dis 2007;66:771–7. https://doi.org/doi:10.1136/ard.2006.060384.
[26] Kapetanovic MC. Axial spondylartrit (ankyloserande spondylit, pelvospondylit, Bechterews sjukdom) 2025. https://www.internetmedicin.se/reumatologi/axial-spondylartrit-ankyloserande-spondylit-pelvospondylit-bechterews-sjukdom (accessed March 30, 2026).
[27] Bruce B, Fries JF. The Stanford Health Assessment Questionnaire: dimensions and practical applications. Health Qual Life Outcomes 2003;1:20. https://doi.org/10.1186/1477-7525-1-20.
[28] Thyberg I, Dahlström Ö, Björk M, Arvidsson P, Thyberg M. Potential of the HAQ score as clinical indicator suggesting comprehensive multidisciplinary assessments: the Swedish TIRA cohort 8 years after diagnosis of RA. Clin Rheumatol 2012;31:775–83. https://doi.org/10.1007/s10067-012-1937-0.
[29] Fritz CO, Morris PE, Richler JJ. Effect size estimates: current use, calculations, and interpretation. J Exp Psychol Gen 2012;141:2–18. https://doi.org/10.1037/a0024338.
[30] Coolican H. Research methods and statistics in psychology, 6th ed. New York, NY, US: Psychology Press; 2014.
[31] Schoels MM, Aletaha D, Alasti F, Smolen JS. Disease activity in psoriatic arthritis (PsA): defining remission and treatment success using the DAPSA score. Ann Rheum Dis 2016;75:811–8. https://doi.org/10.1136/annrheumdis-2015-207507.
[32] Fransen J, van Riel PLCM. The Disease Activity Score and the EULAR response criteria. Rheum Dis Clin North Am 2009;35:745–57, vii–viii. https://doi.org/10.1016/j.rdc.2009.10.001.

[33] Gülfe A, Geborek P, Saxne T. Response criteria for rheumatoid arthritis in clinical practice: how useful are they? Ann Rheum Dis 2005;64:1186–9. https://doi.org/10.1136/ard.2004.027649.

[34] Janssen MF (Bas), Birnie E, Bonsel GJ. Evaluating the discriminatory power of EQ-5D, HUI2 and HUI3 in a US general population survey using Shannon's indices. Qual Life Res 2007;16:895–904. https://doi.org/10.1007/s11136-006-9160-6.

[35] Pickard AS, De Leon MC, Kohlmann T, Cella D, Rosenbloom S. Psychometric comparison of the standard EQ-5D to a 5 level version in cancer patients. Med Care 2007;45:259–63. https://doi.org/10.1097/01.mlr.0000254515.63841.81.

[36] Janssen MF, Birnie E, Haagsma JA, Bonsel GJ. Comparing the Standard EQ-5D Three-Level System with a Five-Level Version. Value Health 2008;11:275–84. https://doi.org/10.1111/j.1524-4733.2007.00230.x.

[37] Janssen MF, Pickard AS, Golicki D, Gudex C, Niewada M, Scalone L, et al. Measurement properties of the EQ-5D-5L compared to the EQ-5D-3L across eight patient groups: a multi-country study. Qual Life Res 2013;22:1717–27. https://doi.org/10.1007/s11136-012-0322-4.

[38] Devlin NJ, Parkin D, Browne J. Patient-reported outcome measures in the NHS: new methods for analysing and reporting EQ-5D data. Health Econ 2010;19:886–905. https://doi.org/10.1002/hec.1608.

[39] Buchholz I, Thielker K, Feng Y-S, Kupatz P, Kohlmann T. Measuring changes in health over time using the EQ-5D 3L and 5L: a head-to-head comparison of measurement properties and sensitivity to change in a German inpatient rehabilitation sample. Qual Life Res 2015;24:829–35. https://doi.org/10.1007/s11136-014-0838-x.