BMC Rheumatology

RESEARCH

Open Access

# Interchangeability of patient pain, fatigue and global scores in patients with spondyloarthritis - a registry-based simulation study

Stylianos Georgiadis[1*], Daniela Di Giuseppe[2], Almut Scherer[3], Merete Lund Hetland[1,4], Gareth T. Jones[5], Bente Glintborg[1,4,6], Anne Gitte Loft[7,8], Johan K. Wallman[9], Brigitte Michelsen[1,10,11], Eirik Klami Kristianslund[10], Ayten Yazici[12], Merih Birlik[13], Jakub Závada[14,15], Michael J. Nissen[16], Adrian Ciurea[17], Bjorn Gudbjornsson[18,19], Olafur Palsson[9,19], Ziga Rotar[20,21], Matija Tomšič[20,21], Heikki Relas[22], Johanna Huhtakangas[23], Ana Maria Rodrigues[24,25], Maria José Santos[26,27], Isabel Castrejon[28,29], Federico Díaz-González[30,31], Marleen van de Sande[32,33], Pasoon Hellamand[33,34] and Lykke Midtbøll Ørnbjerg[1]

## Abstract

**Background** To investigate a patient-level single imputation approach for patient reported outcomes (PROs) that express similar contents or associated PROs, where a PRO whose value is missing at a particular timepoint is substituted by another PRO whose value is available at the same timepoint.

**Methods** We performed a simulation study on registry-based spondyloarthritis data to explore the potential interchangeability between the patient pain (PPA) and fatigue (PFA) assessment scores and relevant Bath Ankylosing Spondylitis Disease Activity Index (BASDAI) individual questions, and between PPA, PFA and patient global assessment (PGA). Performance was assessed per imputation method in terms of relative bias and coverage. Sample size, level of missingness and missing data pattern were included as parameters in the simulations.

**Results** All applied scenarios to interchange PPA with BASDAI question 2 (axial pain), BASDAI question 3 (peripheral joint pain/swelling) or their average failed. Interchangeability between PFA and BASDAI question 1 (fatigue/tiredness) was acceptable for partially (up to 50%) missing data. When interchanging patient assessment scores (PPA, PFA and PGA), we observed inconsistent results in terms of performance. The performance of the applied methods depended on the sample size and the level of missingness, but not heavily on the underlying missing data pattern.

**Conclusions** Interchanging PFA and the BASDAI fatigue question was justified for partially missing data, while interchangeability between PPA, PFA and PGA, and between PPA and the BASDAI pain questions was not advised. Our findings suggest that registering patient assessment scores and BASDAI questions is recommended.

*Correspondence:
Stylianos Georgiadis
stylianos.georgiadis@regionh.dk

Full list of author information is available at the end of the article

**Keywords** Spondyloarthritis, Patient reported outcomes, Missing data, Single imputation, Registry data

## Background

Patient reported outcomes (PROs) are key features in the clinical assessment of patients with spondyloarthritis, including axial spondyloarthritis (axSpA) and psoriatic arthritis (PsA) [1–4]. The single-item PROs of patient pain assessment (PPA), patient fatigue assessment (PFA) and patient global assessment (PGA) are frequently used in both diagnoses [5, 6]. Besides, multi-item PROs combining several questions on patient's disease activity and related health status have been developed, such as the Bath Ankylosing Spondylitis Disease Activity Index (BASDAI), a 6-item questionnaire on fatigue, axial pain, peripheral joint pain/swelling, discomfort and morning stiffness typically used in axSpA [7]. PRO items are also part of most composite outcome scores along with physician reported measures and laboratory values, e.g., the Axial Spondyloarthritis Disease Activity Score (ASDAS) that includes PGA and questions 2, 3 and 6 from BASDAI [8], and the Disease Activity index for PSoriatic Arthritis (DAPSA) which includes PPA and PGA [9, 10].

Missing data is a challenge in all types of studies, including clinical trials [11] and epidemiological studies [12]. This issue may be particularly prevalent in observational studies with data collected in clinical practice, where reasons for missing data may include: missing visits, patients lost to follow-up, a variable not registered at a certain time or center, or lack of resources to ensure data collection in clinical care. Additionally, in multicohort collaboration studies, a variable of interest may be missing in one or some of these cohorts. In a real-world setting, missing data are usually not random but depend on other observed variables [13].

Several approaches for the handling of missing data have been proposed for observational studies [14, 15]. Complete case analysis (CCA) is a common approach in clinical research [16]. A popular alternative is multiple imputation which is considered to be an improvement over CCA [17, 18]. Another approach to handle missing data is single imputation and various single imputation methods, such as last observation carried forward, have been suggested [19]. Single imputation may be preferred over CCA because no observations are sacrificed, but this approach performs poorer than multiple imputation [17]. Multiple imputation, however, can be a difficult method to implement correctly [20], which might make single imputation a valuable alternative. In rheumatology, CCA and multiple imputation are the most frequently methods for dealing with missing observational data [21], while last observation carried forward and non-responder imputation are two commonly used single imputation methods in randomised controlled trials [22, 23].

An unexplored potential option to handle missing data in PROs is to leverage the fact that the same domain is assessed in multiple PROs and that different PROs are strongly associated. For example, the BASDAI evaluates patient's overall level of fatigue/tiredness (BASDAI question 1; Q1) and patient's overall level of axial pain (BASDAI question 2; Q2) and peripheral joint pain/swelling (BASDAI question 3; Q3), meanwhile fatigue and pain are also assessed by PFA and PPA. Moreover, PPA and PFA have been identified as drivers of PGA in rheumatoid arthritis [24, 25], while fatigue has been linked to pain [26]. In light of this, the purpose of this study was to explore a patient-level single imputation approach, where PROs that express similar contents or associated PROs may be potentially interchangeable, i.e., whether one variable whose value is missing at a particular timepoint is substituted by a second variable whose value is available at the same timepoint, and vice versa.

Thus, we conducted a simulation study to investigate the potential interchangeability of patient pain, fatigue and global scores: (i) PPA and the BASDAI pain questions; (ii) PFA and the BASDAI fatigue question; and secondarily (iii) the three single-item patient assessment scores PPA, PFA and PGA. Objective (i) was addressed separately for axSpA and PsA patients, while objectives (ii) and (iii) are axSpA-specific.

## Methods

### EuroSpA data

Analyses relied on anonymised prospectively collected data of patients registered with a diagnosis of axSpA or PsA in 15 European registries participating in the European Spondyloarthritis (EuroSpA) Research Collaboration Network (RCN) [27–29]; ARC (Netherlands), ATTRA (Czech Republic), BIOBADASER (Spain), Biorx.si (Slovenia), BSRBR-AS (United Kingdom), DANBIO (Denmark), GISEA (Italy), ICEBIO (Iceland), NOR-DMARD (Norway), Reuma.pt (Portugal), ROB-FIN (Finland), RRBR (Romania), SCQM (Switzerland), SRQ (Sweden) and TURKBIO (Turkey).

The inclusion criteria were a clinical diagnosis of axSpA or PsA at an age of ≥18 years and the initiation of a tumour necrosis factor inhibitor (TNFi) as first biological treatment, with a start date until December 31st, 2018. Only the first three TNFi treatment series for each patient were included in the data. Data at baseline and three follow-up visits (6, 12 and 24 months) per treatment were collected within specified time windows [27, 28].

The registration of patient assessment scores was not homogeneous across the registries in the EuroSpA

RCN. Questions varied to some degree in their wording, reflecting also different time frames (see Table SA1 in supplementary material A). Furthermore, PROs were registered on different scales (0–10 numeric rating scale versus 0–100 mm visual analogue scale) across and within registries. Since PROs were used in a 0–10 scale in the disease activity scores, we harmonised PRO data to a common standard by rounding 0-100 scales to the nearest 10 and thus converting 0-100 values to a 0–10 scale for all registries.

### Simulation framework

We conducted a simulation study [30, 31] to evaluate the potential interchangeability of patient pain, fatigue and global scores. The conceptual simulation framework used in this study is illustrated in Supplementary figure SA1. We denote the patient assessment scores and the individual BASDAI questions as outcomes of interest for each objective of the study.

### Data-generating mechanism

Simulations were conducted on complete case data for the relevant outcomes (i.e., pooled registrations with complete data for each relevant outcomes of interest at baseline, 6, 12 or 24 months after treatment), where $n_{sim}$ simulated datasets of size $n_{obs}$ were drawn at random. We set $n_{sim} = 1000$. Different sizes $n_{obs}$ for the simulated datasets were applied, i.e., $n_{obs}$ = 100, 200, 500, 1000, 1500 and 2000. Next, for each of the simulated datasets, missing values were introduced at a specific level of missingness (λ), i.e., the proportion of missing data to be created, and according to a chosen missing data pattern. In this simulation study, we varied the level of missingness λ = 10%, 20%, …, 100%.

Three different missing data patterns exist [32]: (a) missing completely at random (MCAR), when the probability of an observation being missing does not depend on any observed and unobserved (missing) variable; (b) missing at random (MAR), when the probability of an observation being missing depends only on the observed variables; and (c) missing not at random (MNAR), when the probability of an observation being missing additionally depends on one or more unobserved variables.

For each outcome of interest, MCAR or MAR data were generated. To generate MCAR data, registrations with missing data were selected from complete case data by random sampling without replacement. Simulating MAR data makes use of all registrations from patients with at least one available registration of any relevant outcome of interest (noted as aim-specific data) and involves the following process. For a given outcome of interest, each registration received a propensity score, i.e., a probability that the given outcome of interest was missing conditional on observed variables [33]. The propensity score was estimated using a logistic regression model on aim-specific data. We considered the following complete variables in the model generating MAR values: gender, age[1], registration year of visit and treatment series number. Registrations with missing data were sampled without replacement and with unequal probabilities (i.e., the propensity scores) via random systematic sampling [34]. As missing data are commonly MAR in real world, rather than MCAR or MNAR [13], results on MAR data were chosen to be primarily presented.

Missing values were then imputed according to the suggested imputation method, i.e., substitution of a missing patient assessment score with one of the remaining scores in the same registration, and a complete dataset with imputed missing values was eventually created. In addition, we assessed CCA as a comparator method.

### Performance

For each parameter of the data-generating mechanism (sample size, level of missingness, missing data pattern and imputation method), the performance of applied methods was assessed. The population parameter of interest was the expected value of the outcomes of interest and the true value of the expected value of the outcomes of interest was determined using the complete case data. We assessed performance of a method in terms of relative bias (to the true value) and coverage of 95% confidence intervals (CIs). CIs based on Monte Carlo standard errors were also calculated to reflect the uncertainty in estimated performance measures.

We chose coverage of 95% CIs (e.g., the proportion of times that the 95% CI contains the true parameter value) as the primary endpoint for this study. By its definition, coverage should be approximately equal to the nominal coverage rate (e.g., 95%) and thus coverage was considered acceptable if Monte Carlo CIs of coverage included the nominal coverage rate. An upper limit of the Monte Carlo CI below 0.95 was an indication of under-coverage for the applied method, while a lower limit of the Monte Carlo CI above 0.95 indicated over-coverage. Before concluding the acceptance of an imputation method, relative bias was also inspected.

## Results

### Patient reported outcomes and data

Patient pain, fatigue and global scores, as expected, had in general higher values at baseline than those at follow-up visits. Descriptive statistics for PROs under consideration and available data per registry and visit in the overall data are given in Supplementary tables SA2a and

[1] Since the exact date of birth was not provided in the data, we assume that: age at 6 months after treatment start = age at baseline; age at 12 months = age at baseline + 1 year; and age at 24 months = age at baseline + 2 years.

SA2b for axSpA and PsA, respectively. Furthermore, PROs displayed similar distributions at the 6-, 12- and 24-month visits across registries (results not shown). Therefore, we performed all simulations on baseline and aggregated follow-up data separately. Table 1 provides an overview of the data used in the analyses per objective. Supplementary figures SA2a-d illustrate the distributions of the PROs per objective at baseline and follow-up visits in complete case data.

### Interchanging patient pain assessment and BASDAI pain questions

We assessed the interchangeability between PPA and the two BASDAI questions about pain (BASDAI Q2 and BASDAI Q3), and we also considered the average of the two BASDAI pain questions as an additional scenario. In terms of coverage and relative bias, none of the two BASDAI questions nor their average had acceptable performance (Fig. 1 and Figure SA3 in supplementary material A). Among the three scenarios, BASDAI Q2 performed best. For missing values at follow-up visits, substituting PPA with the average of the two BASDAI pain questions had an acceptable coverage only with very low levels of missing data, i.e., 10% (Fig. 2.D), and resulted in a fair under-estimation of the relative bias (Figure SA4.D in supplementary material A). Moreover, we evaluated the interchangeability between BASDAI Q2 and Q3, but this scenario performed extremely poorly (Figs. 1.G and J and Figs. 2.G and J).

**Table 1** Overview of data used in the analyses per objective

| Objective | Outcomes of interest | Diagnosis | Number of registries | Registries | Number of patients with at least one available registration | Number of registrations in aim-specific data | Number of registrations in complete case data | Outcomes of interest in complete case data, mean (SD) | |
|---|---|---|---|---|---|---|---|---|---|
| (i) | PPA<br>BASDAI Q2<br>BASDAI Q3 | axSpA | 10 | ATTRA<br>Biorx.si<br>BSRBR-AS<br>DANBIO<br>ICEBIO<br>NOR-DMARD<br>ROB-FIN<br>SCQM<br>SRQ<br>TURKBIO | 20,336 | 25,852 (baseline)<br>46,751 (follow-up) | 15,151 (baseline)<br>31,016 (follow-up) | PPA<br>BASDAI Q2<br>BASDAI Q3<br>PPA<br>BASDAI Q2<br>BASDAI Q3 | 6.0 (2.4)<br>6.4 (2.5)<br>4.6 (3.0)<br>3.2 (2.6)<br>3.5 (2.8)<br>2.6 (2.7) |
| (ii) | PFA<br>BASDAI Q1 | axSpA | 5 | DANBIO<br>ICEBIO<br>NOR-DMARD<br>SRQ<br>TURKBIO | 12,484 | 16,252 (baseline)<br>28,754 (follow-up) | 9,433 (baseline)<br>19,432 (follow-up) | PFA<br>BASDAI Q1<br>PFA<br>BASDAI Q1 | 6.1 (2.7)<br>5.9 (2.7)<br>3.9 (3.0)<br>3.9 (3.0) |
| (iii) | PPA<br>PFA<br>PGA | axSpA | 5 | DANBIO<br>ICEBIO<br>NOR-DMARD<br>SRQ<br>TURKBIO | 13,887 | 18,234 (baseline)<br>31,308 (follow-up) | 12,080 (baseline)<br>23,685 (follow-up) | PPA<br>PFA<br>PGA<br>PPA<br>PFA<br>PGA | 5.9 (2.4)<br>6.1 (2.7)<br>6.1 (2.5)<br>3.2 (2.7)<br>3.9 (3.0)<br>3.5 (2.8) |
| (i) | PPA<br>PFA<br>PGA | PsA | 7 | ATTRA DANBIO<br>ICEBIO<br>NOR-DMARD<br>SCQM<br>SRQ<br>TURKBIO | 11,906 | 15,716 (baseline)<br>25,852 (follow-up) | 9,323 (baseline)<br>16,977 (follow-up) | PPA<br>PFA<br>PGA<br>PPA<br>PFA<br>PGA | 5.9 (2.4)<br>6.1 (2.7)<br>6.1 (2.5)<br>3.5 (2.7)<br>4.2 (3.0)<br>3.8 (2.8) |

Objectives: (i) interchangeability between PPA and BASDAI pain questions; (ii) interchangeability between PFA and BASDAI fatigue question; and (iii) interchangeability between PPA, PFA and PGA. Aim-specific data: registrations from patients with at least one available registration of any relevant outcome of interest. Complete case data: pooled registrations with complete data for each relevant outcomes of interest at baseline, 6, 12 or 24 months after treatment

Phrasing of the individual BASDAI questions:

BASDAI Q1: How would you describe the overall level of fatigue/tiredness you have experienced?

BASDAI Q2: How would you describe the overall level of inflammatory neck, back or hip pain you have had?

BASDAI Q3: How would you describe the overall level of pain/swelling in joints other than neck, back or hips you have had?

axSpA: axial spondyloarthritis; BASDAI: Bath Ankylosing Spondylitis Disease Activity Index; BASDAI Q1: BASDAI question 1; BASDAI Q2: BASDAI question 2; BASDAI Q3: BASDAI question 3; PFA: patient fatigue assessment; PGA: patient global assessment; PPA: patient pain assessment; PsA: psoriatic arthritis; SD: standard deviation

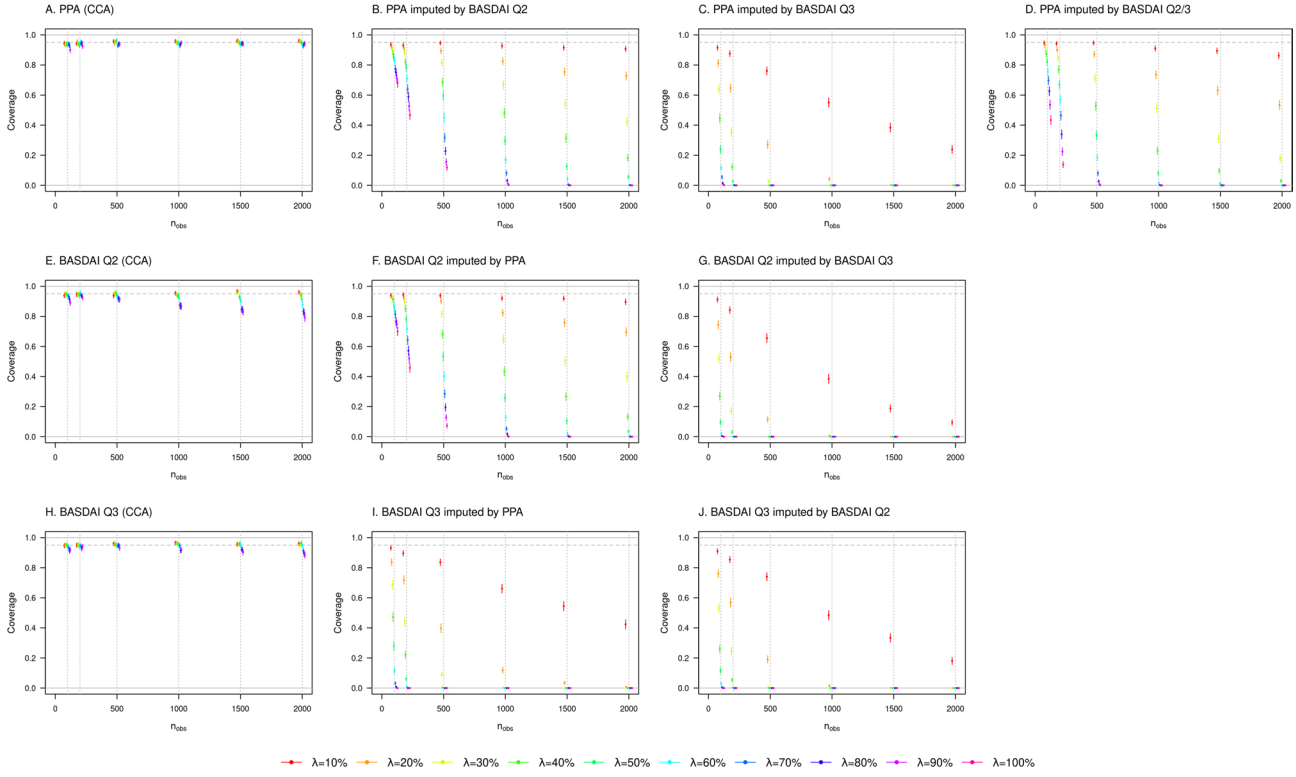


**Fig. 1** Coverage of 95% CIs (points) and corresponding Monte Carlo 95% CIs (line segments) per applied method for patient pain assessment scores (PPA, BASDAI Q2 and BASDAI Q3) for axSpA patients at baseline in MAR data. We varied the sample size $n_{obs}$ = 100, 200, 500, 1000, 1500 and 2000, and the level of missingness λ = 10%, 20%, …, 100%. The number of simulations was $n_{sim} = 1000$. Coverage of 95% CIs is defined as the proportion of times that the 95% CI contains the true parameter value. axSpA: axial spondyloarthritis; BASDAI: Bath Ankylosing Spondylitis Disease Activity Index; BASDAI Q2: BASDAI question 2; BASDAI Q3: BASDAI question 3; BASDAI Q2/3: average of BASDAI Q2 and BASDAI Q3; CCA: complete case analyses; CI: confidence interval; MAR: missing at random; PPA: patient pain assessment

### Interchanging patient fatigue assessment and BASDAI fatigue question

When interchanging PFA and BASDAI Q1, the coverage was acceptable for levels of missingness up to 70% and 50% for missing values at either baseline or follow-up, respectively, while under-coverage was observed above these levels of missingness (Figs. 3 and 4). However, over-coverage was detected for a sample size of 2,000 at baseline. Although the relative bias was very low at all timepoints, the Monte Carlo CIs did not include zero, as seen in Figures SA5 and SA6 in supplementary material A.

### Interchanging patient assessment scores

When interchanging patient assessment scores (PPA, PFA and PGA) in patients with axSpA, we observed inconsistent results between baseline and follow-up visits. The results of our simulations showed that for MAR values at baseline, coverage of 95% CIs when interchanging PFA and PGA was acceptable for levels of missingness up to 60%, but some over-coverage was observed for large sample sizes (Figures SA7a.F and I in supplementary material A). Low levels of bias were observed for interchangeability between PFA and PGA, albeit at a relative bias below 0.01, as can be seen in the Monte Carlo CIs for relative bias that did not include zero (Figures SA7b.F and I in supplementary material A). On the contrary, the performance of the simulations was poor, with a high degree of under-coverage, when interchanging either PPA and PFA or PPA and PGA (Figures SA7a.B and E and Figures SA7a.C and H, respectively), especially for levels of missingness above 20%. When values at follow-up visits were missing, interchangeability between all patient assessment scores performed poorly in terms of coverage and relative bias (Figures SA8a-b in supplementary material A). Among the three pairwise comparisons, interchanging PPA and PGA displayed better results with acceptable performance for a level of missingness of 10% (Figures SA8a.C and H).

When assessing the interchangeability of the patient assessment scores in PsA, no major differences were observed as compared to axSpA (Figures SA9a-b and Figures SA10a-b in supplementary material A).

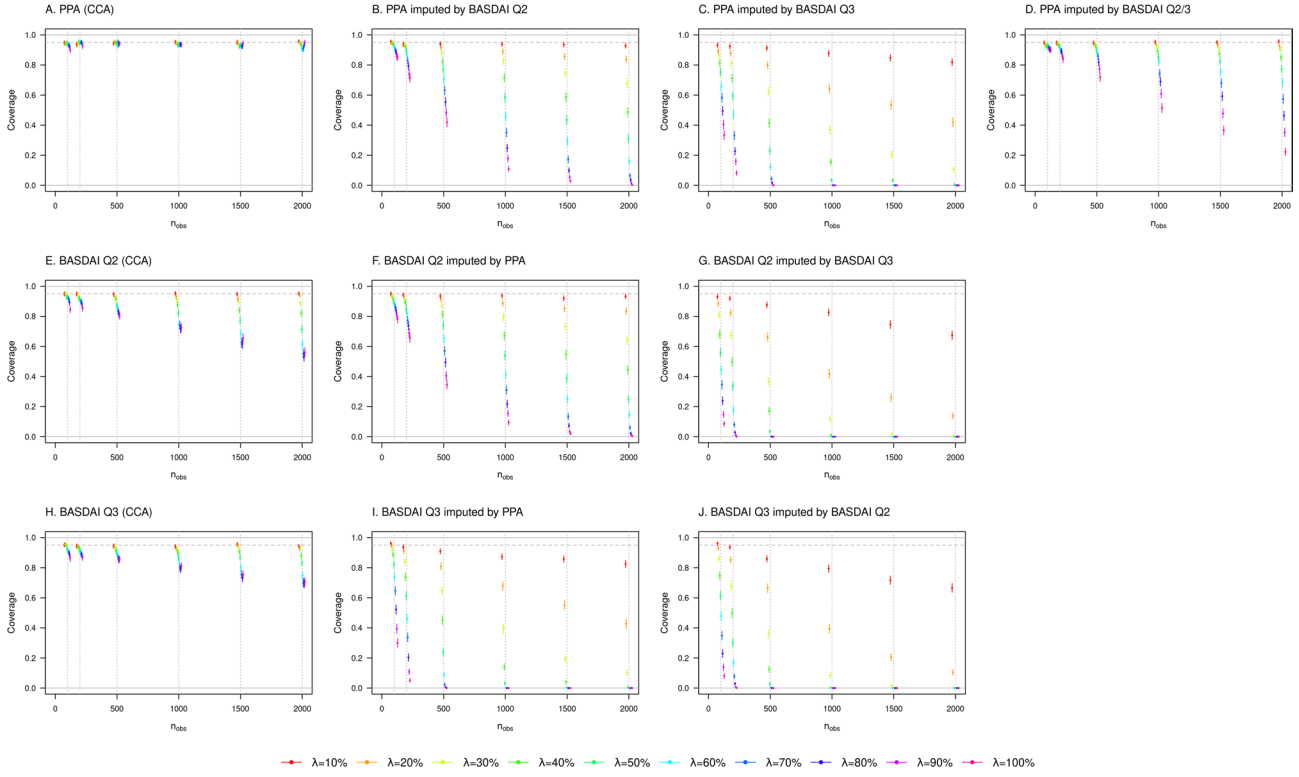


**Fig. 2** Coverage of 95% CIs (points) and corresponding Monte Carlo 95% CIs (line segments) per applied method for patient pain assessment scores (PPA, BASDAI Q2 and BASDAI Q3) for axSpA patients at follow-up in MAR data. We varied the sample size $n_{obs}$ = 100, 200, 500, 1000, 1500 and 2000, and the level of missingness λ = 10%, 20%, …, 100%. The number of simulations was $n_{sim} = 1000$. Coverage of 95% CIs is defined as the proportion of times that the 95% CI contains the true parameter value. axSpA: axial spondyloarthritis; BASDAI: Bath Ankylosing Spondylitis Disease Activity Index; BASDAI Q2: BASDAI question 2; BASDAI Q3: BASDAI question 3; BASDAI Q2/3: average of BASDAI Q2 and BASDAI Q3; CCA: complete case analyses; CI: confidence interval; MAR: missing at random; PPA: patient pain assessment

### Single imputation versus complete case analysis

CCA outperformed interchangeability methods in terms of coverage and relative bias when assessing the patient assessment (PPA, PFA and PGA) and patient pain assessment scores (PPA, BASDAI Q2 and BASDAI Q3). However, for patient fatigue assessment scores (PFA and BASDAI Q1), CCA did not show superior performance.

### Analyses on MCAR data

The MCAR missing data pattern was also assessed (see supplementary material B). The estimates of the performance measures were not drastically affected when the interchangeability approach was applied. In addition, CCA provided unbiased results for MCAR data.

### Impact of imputation in registry data

The potential impact of the interchangeability approach in our data was examined by comparing descriptive statistics before and after imputing the data from patients with at least one available registration of relevant outcomes of interest (Tables SA3a-d in supplementary material A). Among the patient assessment scores (PPA, PFA and PGA), PFA had the lowest availability. The simulation results showed that the mean value of PFA remained consistent when imputed by PGA at baseline, with approximately 25% and 35% missing data in axSpA and PsA patients, respectively. This was also the case when interchanging PFA and BASDAI Q1 at both baseline and follow-up visits.

## Discussion

We conducted a simulation study on real-world data to investigate whether interchanging PROs reflecting similar contents was feasible in axSpA and PsA in order to increase the availability of PROs and derived outcome measures in observational studies. Our findings highlight the complexities of a single imputation approach in PROs and the limitations of interchanging PFA, PPA, PGA, and BASDAI questions.

In the first part of our analyses, we explored the potential interchangeability of patient assessment scores (PPA and PFA) with relevant BASDAI questions (Q1, Q2 and Q3). We demonstrated that PFA and BASDAI Q1 were interchangeable at both baseline and follow-up visits up to an extended missingness (i.e., 50%). Regarding patient pain assessment scores (PPA, BASDAI Q2 and BASDAI

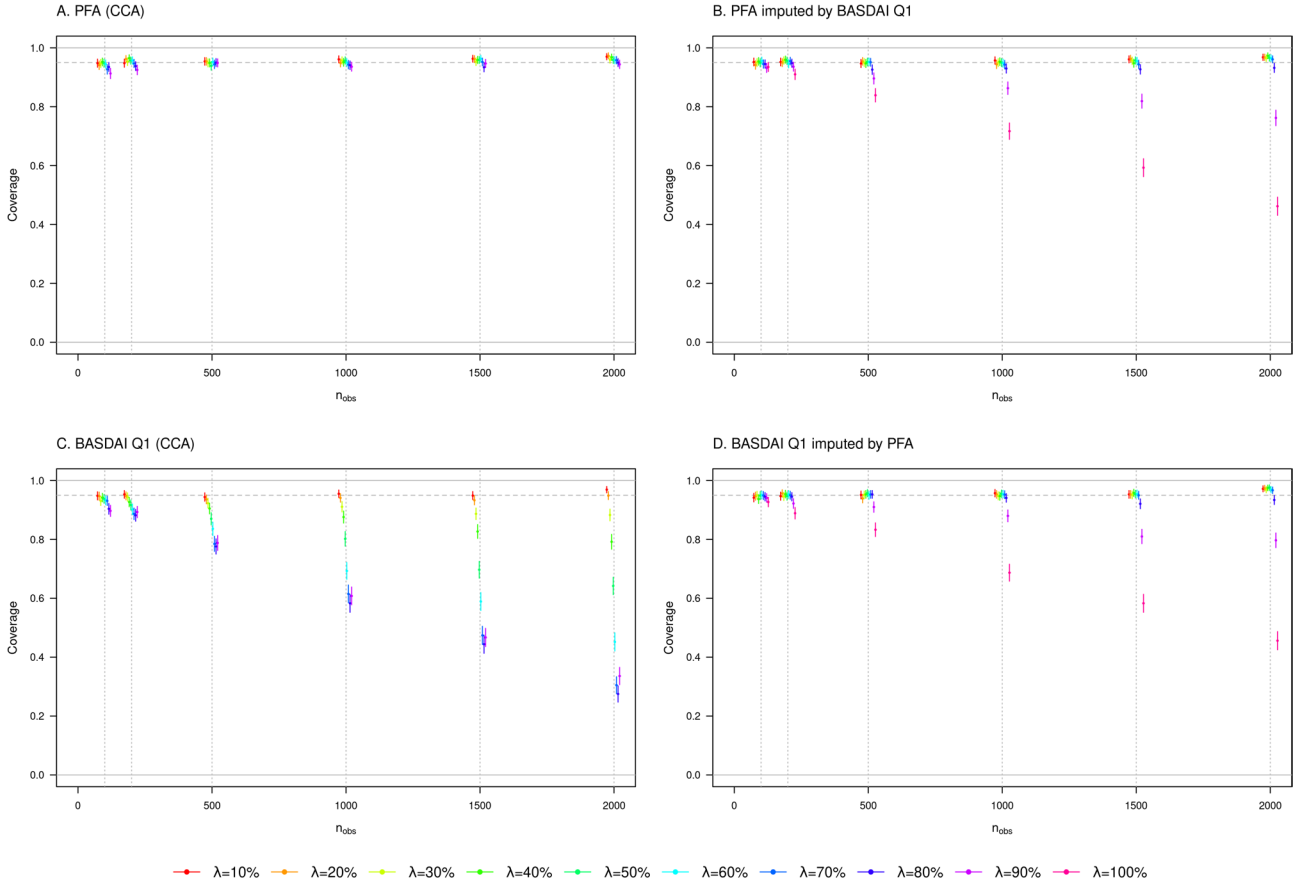


**Fig. 3** Coverage of 95% CIs (points) and corresponding Monte Carlo 95% CIs (line segments) per applied method for patient fatigue assessment scores (PFA and BASDAI Q1) for axSpA patients at baseline in MAR data. We varied the sample size $n_{obs}$ = 100, 200, 500, 1000, 1500 and 2000, and the level of missingness λ = 10%, 20%, …, 100%. The number of simulations was $n_{sim} = 1000$. Coverage of 95% CIs is defined as the proportion of times that the 95% CI contains the true parameter value. axSpA: axial spondyloarthritis; BASDAI: Bath Ankylosing Spondylitis Disease Activity Index; BASDAI Q1: BASDAI question 1; CCA: complete case analyses; CI: confidence interval; MAR: missing at random; PFA: patient fatigue assessment

Q3), results were not consistent between analyses at baseline and follow-up, and none of the interchangeability approaches yielded acceptable performance. Similar results were obtained when interchanging BASDAI Q2 and BASDAI Q3.

In the second part of our analyses, we examined the single-item PPA, PFA and PGA. We found that for missing values at baseline, PFA and PGA were interchangeable in terms of relative bias and coverage for extended levels of missingness (i.e., up to 60%). Nevertheless, when values were missing at follow-up, no method showed an acceptable performance for substantial amounts of missing data. For both axSpA and PsA, we observed similar patterns.

The preceding results were supported by the PRO distributions and the descriptive statistics before and after imputing the relevant data. The main results of our study assumed a MAR missing data pattern, but we also assessed the MCAR pattern. The fact that independent variables in the logistic regression models used for simulating MAR data were significant for all objectives supported the assumption that our data were not MCAR.

We aimed to establish the interchangeability of patient pain, fatigue and global scores that are either strongly associated or assess the same domain. We examined the interchangeability between PPA, PFA and PGA. PGA covers the broad concepts of global health and overall disease activity in clinical practice and is linked to pain and fatigue [35]. Nevertheless, PGA was not found overall interchangeable with either PPA or PFA based on the criteria used in this study. When PROs assess the same domain, PFA was interchangeable with the BASDAI fatigue question only when partially missing (i.e., 50%), but not when one of them was entirely missing (i.e., level of missingness of 100%). In contrast, neither BASDAI pain questions nor their average were interchangeable with PPA at any level of missingness. These findings suggest that in datasets with partially missing information on fatigue assessment it is feasible to impute the missing scores with available scores on BASDAI Q1; or vice versa. However, in the setting of multi-center studies where a

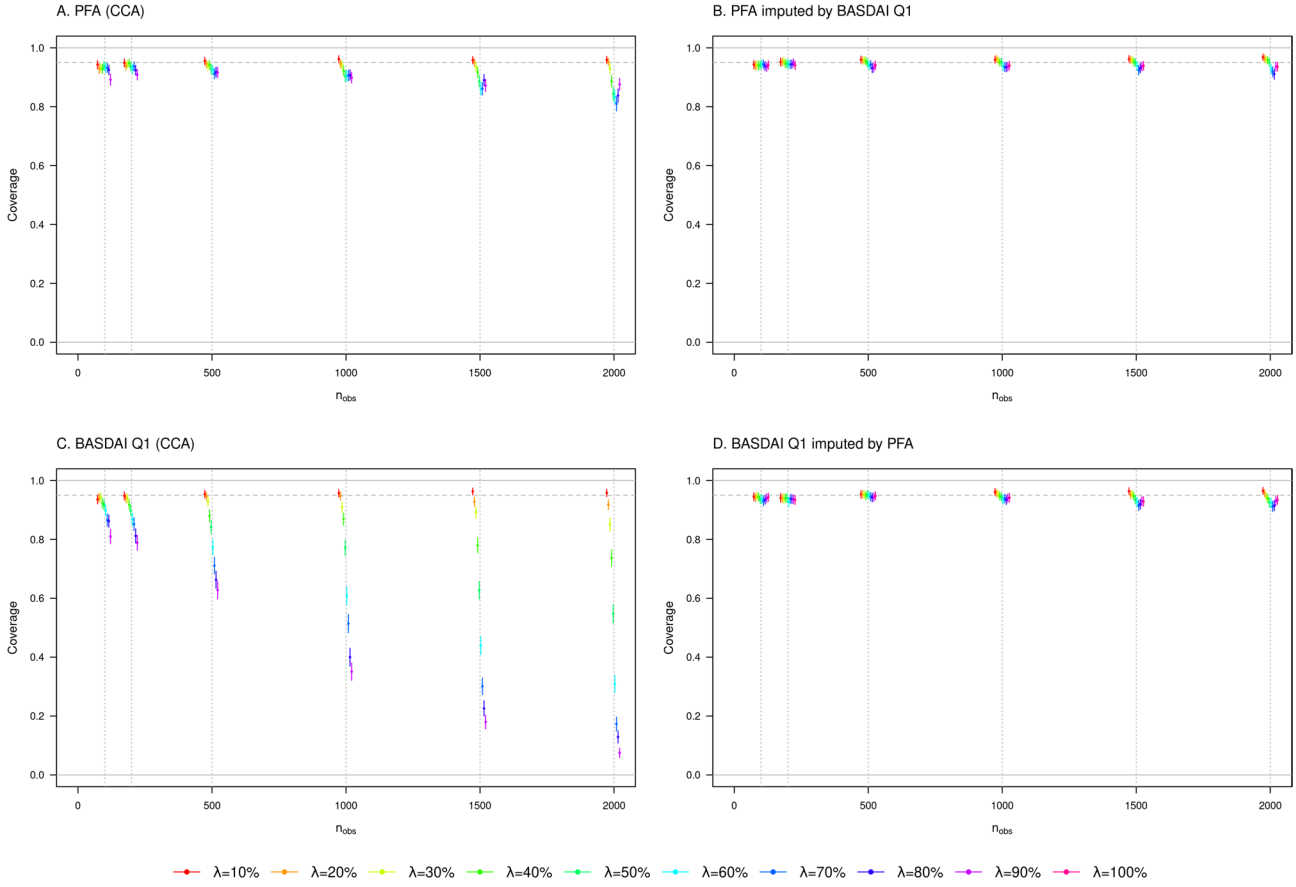


**Fig. 4** Coverage of 95% CIs (points) and corresponding Monte Carlo 95% CIs (line segments) per applied method for patient fatigue assessment scores (PFA and BASDAI Q1) for axSpA patients at follow-up in MAR data. We varied the sample size $n_{obs}$ = 100, 200, 500, 1000, 1500 and 2000, and the level of missingness λ = 10%, 20%, …, 100%. The number of simulations was $n_{sim} = 1000$. Coverage of 95% CIs is defined as the proportion of times that the 95% CI contains the true parameter value. axSpA: axial spondyloarthritis; BASDAI: Bath Ankylosing Spondylitis Disease Activity Index; BASDAI Q1: BASDAI question 1; CCA: complete case analyses; CI: confidence interval; MAR: missing at random; PFA: patient fatigue assessment

single center does not register fatigue or BASDAI, this imputation approach is not valid. Also, interchangeability of PPA and either BASDAI Q2 or Q3 is not recommended based on our chosen endpoint.

A potential reason for the varying results across PROs could be that the domain of fatigue is more specific than the pain domain, leading to greater similarity between fatigue assessments than in pain assessments that consider axial, peripheral and overall pain. Moreover, differences in the phrasing and recall periods of the three assessment scores (PPA, PFA and PGA) in the different countries is expected to influence the interchangeability results in the various scenarios. To promote future multicohort collaborative research, it would be beneficial to work towards harmonization of the phrasing and recall period of patient assessment scores. Overall, our findings support the registration practice of both patient assessment scores and BASDAI questions being collected. Nevertheless, the registration of PFA may be redundant when BASDAI Q1 is collected.

An interesting finding in our analyses was that the baseline patient assessment scores (i.e., PPA, PFA and PGA) showed higher degrees of interchangeability than the follow-up scores. This finding may reflect that high levels of disease activity can affect all domains, while the initiated treatment could impact each domain to a different extent. In line with this, Deodhar et al. [36] showed larger improvements in pain and global assessments than in fatigue assessment in axSpA patients after initiation of secukinumab treatment. Also, we observed the same pattern of interchangeability in both axSpA and PsA patients, supporting the generalizability of our results.

In the context of this study, interchangeability between two variables was defined as the substitution of one variable whose value is missing at a particular timepoint with the second variable whose value was available at the same timepoint (and vice versa). Interchangeability is in fact a patient-level single imputation method. Single imputation approaches may result in bias, but they are considered reasonable and may perform better than CCA [17, 37, 38]. Furthermore, since only one estimate

is calculated, single imputation methods are alluring as compared to other more sophisticated methods, such as multiple imputation. Multiple imputation is known to be superior to single imputation methods regarding performance [13, 17]. However, correct implementation of multiple imputation is challenging, especially when applied under MNAR mechanisms, as standard implementations assume MAR [17, 39]. Hence, single imputation approaches can be of great value.

Compared to the suggested interchangeability methods, CCA performed better in terms of relative bias and coverage for patient assessment scores (PPA, PFA and PGA) and patient pain assessment scores (PPA, BASDAI Q2 and BASDAI Q3). On the contrary, for the interchangeability of PFA and BASDAI Q1, CCA performed worse. Notably, the performance of CCA was affected by the missing data pattern more evidently than the corresponding interchangeability approach. While CCA produced unbiased estimates for MCAR data, this was not always the case for MAR data, as expected [17]. In practice, applying CCA in observational studies with a high level of missingness may be unfeasible or lead to a crucial loss of power.

Analyses relied on real-world data from multiple registries which enhanced the relevance and generalizability of this study. However, the heterogeneity in the wording and the time frame of the patient assessment score questions across registries may have influenced the results. Another limitation of this study was the lack of consideration for MNAR data.

## Conclusions

Interchangeability between PFA and BASDAI Q1 was justified for partially (up to 50%) missing data, which provides researchers with a new option to increase data availability of fatigue assessment scores. However, we did not find consistent results in terms of performance when interchanging patient assessment scores (i.e., PPA, PFA and PGA) between baseline and follow-up visits. Hence, interchangeability between these patient assessment scores was not recommended in either axSpA or PsA patients. Furthermore, all applied scenarios to interchange PPA with BASDAI Q2, BASDAI Q3 or their average failed. Our findings thus support the current registration practice in most clinical registries where both patient assessment scores and BASDAI questions are collected.

### Abbreviations

| | |
|---|---|
| axSpA | Axial SpondyloArthritis |
| ASDAS | Axial Spondyloarthritis Disease Activity Score |
| BASDAI | Bath Ankylosing Spondylitis Disease Activity Index |
| BASDAI Q1 | BASDAI question 1 |
| BASDAI Q2 | BASDAI question 2 |
| BASDAI Q3 | BASDAI question 3 |
| CCA | Complete Case Analysis |
| CI | Confidence Interval |
| DAPSA | Disease Activity index for PSoriatic Arthritis |
| EuroSpA | European Spondyloarthritis |
| MAR | Missing At Random |
| MCAR | Missing Completely At Random |
| MNAR | Missing Not At Random |
| PFA | Patient Fatigue Assessment |
| PGA | Patient Global Assessment |
| PPA | Patient Pain Assessment |
| PRO | Patient Reported Outcome |
| PsA | Psoriatic Arthritis |
| RCN | Research Collaboration Network |
| TNFi | Tumour Necrosis Factor inhibitor |

## Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s41927-025-00527-6.

Supplementary Material 1

Supplementary Material 2

### Acknowledgements

The EuroSpA collaboration has been supported by Novartis Pharma AG since 2017 and UCB Biopharma SRL since 2022. This EuroSpA study was financially supported by UCB. No financial sponsors had any influence on the data collection, statistical analyses, manuscript preparation, or decision to submit.

### Author contributions

SG conceived of the original idea. SG an LØ designed the analyses. SG performed the simulations and analysed the data. SG drafted the manuscript with input from DDG, AS, MHL, GTJ and LØ. LØ supervised the project. DDG, AS, MHL, GTJ, BGl, AGL, JKW, BM, EKK, AY, MB, JZ, MJN, AC, BGu, OP, ZR, MT, HR, JH, AMR, MJS, IC, FDG, MGHvdS and PH contributed to the acquisition of the data. All authors contributed to the interpretation of the data. All authors revised critically the manuscript, approved the final version of the manuscript, and agreed with its content. All authors are responsible for the integrity and accuracy of the work.

### Funding

Open access funding provided by Copenhagen University. Open access funding provided by Copenhagen University. Open access funding provided by Copenhagen University. This work was supported by Novartis Pharma AG. Novartis had no influence on the data collection, statistical analyses, manuscript preparation or decision to submit the manuscript.

### Data availability

The data in this article was collected in the individual registries and made available for secondary use through the EuroSpA Research Collaboration Network [https://eurospa.eu/#registries]. Relevant patient level data may be made available on reasonable request to the corresponding author but will require approval from all contributing registries.

## Declarations

### Consent for publication

Not applicable.

### Human ethics and consent to participate

All patient data were collected in accordance with national legal and regulatory requirements in the different countries including obtainment of written informed consent from participating patients if required in the individual country (Table SA4 in supplementary material A). The study was approved by the respective national Data Protection Agencies or Ethical Committees according to legal regulatory requirements in the participating countries. The study was carried out in accordance with the Declaration of Helsinki.

**Competing interests**
Stylianos Georgiadis: Novartis, UCB; Daniela Di Giuseppe: none; Almut Scherer: none; Merete Lund Hetland: Abbvie, Biogen, BMS, CellTrion, Eli Lilly, Janssen Biologics B.V, Lundbeck Fonden, MSD, Medac, Nordforsk, Novartis, Pfizer, Roche, Samsung Bioepis, Sandoz, UCB; Gareth T. Jones: AbbVie, Amgen, GSK, Janssen, Menarini, Pfizer, Shionogi, UCB; Bente Glintborg: Abbvie, BMS, Pfizer, Sandoz; Anne Gitte Loft: Janssen, Novartis, UCB; Johan K. Wallman: AbbVie, AstraZeneca, BMS, Eli Lilly, Janssen, MSD, Pfizer, Roche, Samsung Bioepis, Sanofi, UCB; Brigitte Michelsen: Novartis, Pfizer; Eirik Klami Kristianslund: none; Ayten Yazici: Abbvie, Amgen, Novartis, Pfizer, Roche, UCB; Merih Birlik: Boehringer- Ingelheim; Jakub Závada: Abbvie, AstraZeneca, Egis, Elli Lilly, Glaxo, Novartis, Sandoz, Sanofi, Sobi, Swix-biopharma, UCB; Michael J. Nissen: AbbVie, Amgen, Eli Lilly, Janssens, Novartis, Pfizer; Adrian Ciurea: none; Bjorn Gudbjornsson: none; Olafur Palsson: none; Ziga Rotar: Abbvie, Amgen, AstraZeneca, Biogen, Eli Lilly, Janssen, Lek, Medis, MSD, Novartis, Pfizer, Sanofi, Sobi, Swixx BioPharma; Matija Tomsic: Abbvie, Amgen, Biogen, Eli Lilly, Janssen, Medis, MSD, Novartis, Pfizer, Sandoz-Lek, Sanofi; Heikki Relas: Abbvie, Celgene, Pfizer, UCB, Viatris; Johanna Huhtakangas: Abbvie, Amgen, Boehringer-Ingelheim, Novartis; Ana Maria Rodrigues: Abbvie, Amgen, Novartis, Pfizer; Maria José Santos: Abbvie, Amgen, AstraZeneca, Janssen, Eli Lilly, Medac, Novartis, Pfizer; Isabel Castrejon: Alfasigma, BMS, Boehringer-Ingelheim, Eli Lilly, Galapagos, Gebro, GSK, Janssen, Pfizer, UCB; Federico Díaz-González: Abbvie, BMS, Galapagos, Janssen, MSD, UCB; Marleen van de Sande: Abbvie, Eli Lilly, Janssen, Novartis, UCB; Pasoon Hellamand: Novartis; Lykke Midtbøll Ørnbjerg: Novartis, UCB.

**Author details**
[1]Copenhagen Center for Arthritis Research (COPECARE), Center for Rheumatology and Spine Diseases, Centre for Head and Orthopaedics, Rigshospitalet, Glostrup, Denmark
[2]Clinical Epidemiology Division, Department of Medicine Solna, Karolinska Institutet, Stockholm, Sweden
[3]Statistics Group, Swiss Clinical Quality Management Foundation, Zurich, Switzerland
[4]Department of Clinical Medicine, University of Copenhagen, Copenhagen, Denmark
[5]Aberdeen Centre for Arthritis and Musculoskeletal Health (Epidemiology Group), University of Aberdeen, Aberdeen, UK
[6]The DANBIO Registry, Copenhagen, Denmark
[7]Department of Rheumatology, Aarhus University Hospital, Aarhus, Denmark
[8]Department of Clinical Medicine, Aarhus University, Aarhus, Denmark
[9]Department of Clinical Sciences Lund, Rheumatology, Skåne University Hospital, Lund University, Lund, Sweden
[10]Center for Treatment of Rheumatic and Musculoskeletal Diseases (REMEDY), Diakonhjemmet Hospital, Oslo, Norway
[11]Research Unit, Sørlandet Hospital, Kristiansand, Norway
[12]Division of Rheumatology, Department of Internal Medicine, School of Medicine, Kocaeli University, Kocaeli, Turkey
[13]Division of Rheumatology, Department of Internal Medicine, Dokuz Eylul University School of Medicine, Izmir, Turkey
[14]Institute of Rheumatology, Prague, Czech Republic
[15]Department of Rheumatology, First Faculty of Medicine, Charles University, Prague, Czech Republic
[16]Department of Rheumatology, Geneva University Hospital, Geneva, Switzerland
[17]Department of Rheumatology, University Hospital Zurich, University of Zurich, Zurich, Switzerland
[18]Centre for Rheumatology Research, Landspitali University Hospital, Reykjavik, Iceland
[19]Faculty of Medicine, University of Iceland, Reykjavik, Iceland
[20]Department of Rheumatology, University Medical Centre Ljubljana, Ljubljana, Slovenia
[21]Faculty of Medicine, University of Ljubljana, Ljubljana, Slovenia
[22]Rheumatology, Inflammation Center, Helsinki University Hospital, University of Helsinki, Helsinki, Finland
[23]Department of Rheumatology, Kuopio University Hospital, Kuopio, Finland
[24]EpiDoC Unit, CEDOC, Nova Medical School, Lisbon, Portugal
[25]Rheumatology Unit, Hospital dos Lusíadas, Lisbon, Portugal
[26]Department of Rheumatology, Hospital Garcia de Orta, Almada, Portugal
[27]Instituto Medicina Molecular, Faculdade de Medicina da Universidade de Lisboa, Centro Académico de Medicina de Lisboa, Lisbon, Portugal
[28]Department of Rheumatology, Hospital General, Universitario Gregorio Marañón, Madrid, Spain
[29]Faculty of Medicine, Complutense University of Madrid, Madrid, Spain
[30]Department of Internal Medicine, Dermatology and Psychiatry, Universidad de La Laguna, La Laguna, Spain
[31]Rheumatology Service, Hospital Universitario de Canarias, La Laguna, Spain
[32]Department of Rheumatology & Clinical Immunology and Department of Experimental Immunology, Amsterdam UMC, Amsterdam Institute for Infection & Immunity, University of Amsterdam, Amsterdam, Netherlands
[33]Amsterdam Rheumatology Immunology Center, Reade and Amsterdam UMC, Amsterdam, Netherlands
[34]Department of Rheumatology and Clinical Immunology, Amsterdam University Medical Center, Amsterdam, Netherlands

Received: 5 December 2024 / Accepted: 3 June 2025
Published online: 01 July 2025

**References**

1. Nhan DT, Caplan L. Patient-Reported outcomes in axial spondyloarthritis. Rheumatic Disease Clin North Am. 2016;42(2):285–99.
2. Orbai AM, Ogdie A. Patient-Reported outcomes in psoriatic arthritis. Rheumatic Disease Clin North Am. 2016;42(2):265–83.
3. Png K, Kwan YH, Leung YY, Phang JK, Lau JQ, Lim KK, et al. Measurement properties of patient reported outcome measures for spondyloarthritis: A systematic review. Semin Arthritis Rheum. 2018;48(2):274–82.
4. Torre-Alonso JC, Queiro R, Comellas M, Lizán L, Blanch C. Patient-reported outcomes in European spondyloarthritis patients: a systematic review of the literature. Patient Prefer Adherence. 2018;12:733–47.
5. Sieper J, Rudwaleit M, Baraliakos X, Brandt J, Braun J, Burgos-Vargas R, et al. The assessment of spondyloarthritis international society (ASAS) handbook: A guide to assess spondyloarthritis. Ann Rheum Dis. 2009;68:ii1–44.
6. Orbai AM, De Wit M, Mease P, Shea JA, Gossec L, Leung YY, et al. International patient and physician consensus on a psoriatic arthritis core outcome set for clinical trials. Ann Rheum Dis. 2017;76(4):673–80.
7. Garrett S, Jenkinson T, Kennedy LG, Whitelock H, Gaisford P, Calin A. A new approach to defining disease status in ankylosing spondylitis: the bath ankylosing spondylitis disease activity index. J Rheumatol. 1994;21(12):2286–91.
8. 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(1):18–24.
9. 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(8):1441–7.
10. van der Heijde D, Molto A, Ramiro S, Braun J, Dougados M, van Gaalen F, et al. Goodbye to the term 'ankylosing spondylitis', hello 'axial spondyloarthritis': time to embrace the ASAS-defined nomenclature. Ann Rheum Dis. 2024;83:547–9.
11. Calvert M, Blazeby J, Altman DG, Revicki DA, Moher D, Brundage MD. Reporting of patient-reported outcomes in randomized trials: the CONSORT PRO extension. JAMA - J Am Med Association. 2013;309(8):814–22.
12. Bell ML, Fairclough DL. Practical and statistical issues in missing data for longitudinal patient-reported outcomes. Stat Methods Med Res. 2014;23(5):440–59.
13. Donders ART, van der Heijden GJMG, Stijnen T, Moons KGM. Review: A gentle introduction to imputation of missing values. J Clin Epidemiol. 2006;59(10):1087–91.
14. Pedersen AB, Mikkelsen EM, Cronin-Fenton D, Kristensen NR, Pham TM, Pedersen L, et al. Missing data and multiple imputation in clinical epidemiological research. Clin Epidemiol. 2017;9:157–66.
15. Lee KJ, Tilling KM, Cornish RP, Little RJA, Bell ML, Goetghebeur E, et al. Framework for the treatment and reporting of missing data in observational studies: the treatment and reporting of missing data in observational studies framework. J Clin Epidemiol. 2021;134:79–88.
16. Eekhout I, De Boer RM, Twisk JWR, De Vet HCW, Heymans MW. Missing data: A systematic review of how they are reported and handled. Epidemiology. 2012;23(5):729–32.

17. Schafer JL, Graham JW. Missing data: our view of the state of the Art. Psychol Methods. 2002;7(2):147–77.
18. Hayati Rezvan P, Lee KJ, Simpson JA. The rise of multiple imputation: A review of the reporting and implementation of the method in medical research data collection, quality, and reporting. BMC Med Res Methodol. 2015;15(1):1–14.
19. Engels JM, Diehr P. Imputation of missing longitudinal data: A comparison of methods. J Clin Epidemiol. 2003;56(10):968–76.
20. Sterne JAC, White IR, Carlin JB, Spratt M, Royston P, Kenward MG, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. BMJ (Online). 2009;339(7713):157–60.
21. Lauper K, Kedra J, De Wit M, Fautrel B, Frisell T, Hyrich KL, et al. Analysing and reporting of observational data: A systematic review informing the EULAR points to consider when analysing and reporting comparative effectiveness research with observational data in rheumatology. RMD Open. 2021;7(3):1–9.
22. Ibrahim F, Tom BDM, Scott DL, Prevost AT. A systematic review of randomised controlled trials in rheumatoid arthritis: the reporting and handling of missing data in composite outcomes. Trials. 2016;17:272.
23. Khan NA, Torralba KD, Aslam F. Missing data in randomised controlled trials of rheumatoid arthritis drug therapy are substantial and handled inappropriately. RMD Open. 2021;7:e001708.
24. Khan NA, Spencer HJ, Abda E, Aggarwal A, Alten R, Ancuta C, et al. Determinants of discordance in patients' and physicians' rating of rheumatoid arthritis disease activity. Arthritis Care Res (Hoboken). 2012;64(2):206–14.
25. Desthieux C, Hermet A, Granger B, Fautrel B, Gossec L. Patient-Physician discordance in global assessment in rheumatoid arthritis: A systematic literature review with Meta-Analysis. Arthritis Care Res (Hoboken). 2016;68(12):1767–73.
26. Pollard LC, Choy EH, Gonzalez J, Khoshaba B, Scott DL. Fatigue in rheumatoid arthritis reflects pain, not disease activity. Rheumatology. 2006;45(7):885–9.
27. Ørnbjerg LM, Brahe CH, Askling J, Ciurea A, Mann H, Onen F, et al. Treatment response and drug retention rates in 24 195 biologic-naïve patients with axial spondyloarthritis initiating TNFi treatment: routine care data from 12 registries in the EuroSpA collaboration. Ann Rheum Dis. 2019;78(11):1536–44.
28. Brahe CH, Ørnbjerg LM, Jacobsson L, Nissen MJ, Kristianslund EK, Mann H, et al. Retention and response rates in 14 261 PsA patients starting TNF inhibitor treatment - Results from 12 countries in EuroSpA. Rheumatology. 2020;59(7):1640–50.
29. Christiansen SN, Ørnbjerg LM, Rasmussen SH, Loft AG, Askling J, Iannone F et al. European bio-naïve spondyloarthritis patients initiating TNF inhibitor: time trends in baseline characteristics, treatment retention and response. Rheumatology. 2021.
30. Burton A, Altman DG, Royston P, Holder RL. The design of simulation studies in medical statistics. Stat Med. 2006;25(24):4279–92.
31. Morris TP, White IR, Crowther MJ. Using simulation studies to evaluate statistical methods. Stat Med. 2019;38(11):2074–102.
32. Rubin DB. Inference and missing data. Biometrika. 1976;63(3):581–92.
33. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. Biometrika. 1983;70(1):41–55.
34. Tillé Y. Sampling and Estimation from finite populations. Hoboken, NJ: Wiley; 2020.
35. Nikiphorou E, Radner H, Chatzidionysiou K, Desthieux C, Zabalan C, van Eijk-Hustings Y, et al. Patient global assessment in measuring disease activity in rheumatoid arthritis: A review of the literature. Arthritis Res Ther. 2016;18:251.
36. Deodhar AA, Mease PJ, Rahman P, Navarro-Compán V, Strand V, Hunter T et al. Ixekizumab improves spinal pain, function, fatigue, stiffness, and sleep in radiographic axial spondyloarthritis: COAST-V/W 52-week results. BMC Rheumatol. 2021;5(35).
37. van der Heijden GJMG, Donders T, Stijnen AR, Moons T. Imputation of missing values is superior to complete case analysis and the missing-indicator method in multivariable diagnostic research: A clinical example. J Clin Epidemiol. 2006;59(10):1102–9.
38. Ambler G, Omar RZ, Royston P. A comparison of imputation techniques for handling missing predictor values in a risk model with a binary outcome. Stat Methods Med Res. 2007;16(3):277–98.
39. White IR, Royston P, Wood AM. Multiple imputation using chained equations: issues and guidance for practice. Stat Med. 2011;30(4):377–99.

## Publisher's note