Polysopoulos *et al.*
BMC Medical Research Methodology (2025) 25:55
https://doi.org/10.1186/s12874-025-02515-3

BMC Medical Research Methodology

RESEARCH

Open Access

Handling of missing component information for common composite score outcomes used in axial spondyloarthritis research when complete-case analysis is unbiased

Christos Polysopoulos[1], Stylianos Georgiadis[2], Lykke Midtbøll Ørnbjerg[2], Almut Scherer[1], Daniela Di Giuseppe[3], Merete Lund Hetland[4], Michael John Nissen[5], Gareth T. Jones[6], Bente Glintborg[4], Anne Gitte Loft[7], Johan Karlsson Wallman[8], Karel Pavelka[9], Jakub Závada[9], Ayten Yazici[10], Maria José Santos[11], Adrian Ciurea[12], Burkhard Möller[13], Brigitte Michelsen[14], Pawel Mielnik[15], Johanna Huhtakangas[16], Heikki Relas[17], Katja Perdan Pirkmajer[18], Ziga Rotar[18], Ross MacDonald[6], Bjorn Gudbjornsson[19,20], Irene van der Horst-Bruinsma[21], Marleen van de Sande[22] and Myriam Riek[1*]

Abstract

Background Observational data on composite scores often comes with missing component information. When a complete-case (CC) analysis of composite scores is unbiased, preferable approaches of dealing with missing component information should also be unbiased and provide a more precise estimate. We assessed the performance of several methods compared to CC analysis in estimating the means of common composite scores used in axial spondyloarthritis research.

Methods Individual mean imputation (IMI), the modified formula method (MF), overall mean imputation (OMI), and multiple imputation of missing component values (MI) were assessed either analytically or by means of simulations from available data collected across Europe. Their performance in estimating the means of the Bath Ankylosing Spondylitis Disease Activity Index (BASDAI), the Bath Ankylosing Spondylitis Functional Index (BASFI), and the Ankylosing Spondylitis Disease Activity Score based on C-reactive protein (ASDAS-CRP) in cases where component information was set missing completely at random was compared to the CC approach based on bias, variance, and coverage.

Results Like the MF method, IMI uses a modified formula for observations with missing components resulting in modified composite scores. In the case of an unbiased CC approach, these two methods yielded representative samples of the distribution arising from a mixture of the original and modified composite scores, which, however, could not be considered the same as the distribution of the original score. The IMI and MF method are, thus, intrinsically biased. OMI provided an unbiased mean but displayed a complex dependence structure among observations that, if not accounted for, resulted in severe coverage issues. MI improved precision compared to CC and gave unbiased means and proper coverage as long as the extent of missingness was not too large.

*Correspondence:
Myriam Riek
myriam.riek@scqm.ch
Full list of author information is available at the end of the article

Conclusions MI of missing component values was the only method found successful in retaining CC's unbiasedness and in providing increased precision for estimating the means of BASDAI, BASFI, and ASDAS-CRP. However, since MI is susceptible to incorrect implementation and its performance may become questionable with increasing missingness, we consider the implementation of an error-free CC approach a valid and valuable option.

Trial registration Not applicable as study uses data from patient registries.

Keywords Composite score, Missing components, Complete-case analysis, Multiple imputation, Axial spondyloarthritis

Background

Validated composite scores derived from several components, each measuring different aspects of a patient's disease status, are important tools when aiming to capture the multi-faceted nature of many diseases. For axial spondyloarthritis (axSpA), a chronic inflammatory rheumatic disease primarily affecting the spine, composite scores measuring disease activity and physical function are important tools used in both randomized clinical trials and routine care [1–6].

Composite score values cannot be calculated if one or more component values are missing. Researchers faced with missing component values in their study's sample must therefore take a decision about how to handle the situation. The easiest approach is to disregard observations with incomplete component information and to perform a complete-case (CC) analysis to estimate the distributional parameter of interest. However, this comes with two disadvantages: reduced precision compared to what would have been achievable had all observations been fully observed and the possibility of bias. Understandably, researchers would prefer methods that are able to make use of the available partial information to minimise loss of precision as well as to prevent possible bias.

In this study, we aimed to assess and compare the performance of several alternatives to the CC approach in estimating distributional parameters for commonly used composite score outcomes in axSpA research: the Bath Ankylosing Spondylitis Disease Activity Index [1], BASDAI, the Bath Ankylosing Spondylitis Functional Index [2], BASFI, and the Ankylosing Spondylitis Disease Activity Score based on C-reactive protein (CRP) [3, 4], ASDAS-CRP. These composite scores differ in terms of calculation (a simple arithmetic mean of components in case of BASFI versus a weighted sum of possibly transformed components for ASDAS-CRP) as well as the source (patient or physician reported) and nature (visual analogue scale or numeric rating scale versus laboratory measurement) of the components (Table 1). Composite scores whose components measure different characteristics of a disease on the same scale (with larger values being indicative of a more severe disease), such as the BASDAI and BASFI, are often either weighted or unweighted averages of the components. For such composite scores, researchers are inclined to assume that the arithmetic mean of the observed components provides a reasonable substitute for the missing value of other components [7]. For example, for the BASFI such an individual mean imputation (IMI) approach would simply replace a single missing component value by the arithmetic mean of the remaining nine observed components. Alternatively, instead of imputing a missing component value, that component can simply be discarded from the composite score formula and the composite score derived as the average of the observed components with modified weights that preserve the relations [7, 8]. Regarding the BASFI, this modified composite score formula (MF) approach would simply calculate the arithmetic mean of the observed components. For both approaches, the original sample size is retained and no information is discarded. It is, however, unclear whether these

Table 1 Composite scores

Composite score	Formula[a]
BASDAI	$0.2*(Q_{BASDAI}1+Q_{BASDAI}2+Q_{BASDAI}3+Q_{BASDAI}4)+0.1*(Q_{BASDAI}5+Q_{BASDAI}6)$
BASFI	$0.1*(Q_{BASFI}1+Q_{BASFI}2+Q_{BASFI}3+Q_{BASFI}4+Q_{BASFI}5+Q_{BASFI}6+Q_{BASFI}7+Q_{BASFI}8+Q_{BASFI}9+Q_{BASFI}10)$
ASDAS-CRP	$0.121*[Q_{BASDAI}2]+0.110*$[patient global assessment of disease]$+0.073*[Q_{BASDAI}3]+0.058*[Q_{BASDAI}6]+0.579*[\ln(\max(CRP,2)+1)]$

[a] All patient reported components, i.e., all components but CRP, are assessed by means of visual analogue scales or numerical rating scales from 0 to 10. For detailed information on components, we refer to [1–4] as well as Additional file 1

approaches are unbiased and to what extent they preserve precision compared to a CC analysis.

We can also take advantage of the availability of the other observations in the sample. For instance, replacing a missing component value by the mean of all observed values for that component in the sample (or a suitable subset thereof) is justifiable as a means to replace a missing value by its estimated expected value. This preserves the expected value of the composite score but not its distribution. Such an overall mean imputation (OMI) approach has received criticism in the past [7, 9]. We have included this method in our study nonetheless to investigate in more detail why it fails. Another, relatively new approach is multiple imputation (MI), which, as the name suggests, imputes a missing value not just once but multiple times based on what is learned about the component from the sample [10]. MI methods are now widely used to deal with missing information, particularly missing information in covariates. But they have also previously been assessed in the context of composite scores as the outcome variable such as the EuroQoL 5-Dimensions 3-level utility score, the 12-item Short Form Survey score, and the American College of Rheumatology 20% improvement measure [11–13]. These studies focused on comparing MI at the level of the composite score with MI performed at the level of the components of the score. They found support for a superior performance of MI at the level of the components in terms of precision when partial information on components was available. Such findings make sense intuitively as MI at the composite score level ignores any partially available component information. Moreover, MI at the level of the composite score outcome is considered to have limited potential to outperform a CC approach in terms of precision [10, p. 25][14, 15]. It is therefore reasonable to hypothesize that MI at the component level would outperform CC analysis in this regard.

For this study, we wanted to estimate the expected value, i.e., the population mean, of the composite scores BASDAI, BASFI, and ASDAS-CRP and investigate situations for which analyses restricted to the observations with complete component information (CC approach) are unbiased but suffer from loss of precision due to the reduced sample size. The handling of missing component information in such situations should ideally save the available information with respect to the composite score and obtain a more precise estimation compared to the CC approach, without resulting in bias. The performance of the imputation methods was assessed by means of simulations from available data [16] (for OMI and MI) or analytically (for IMI and MF).

Methods

Data provenance and composite scores of interest

The EuroSpA Research Collaboration Network (RCN) is a registry-based initiative to collaboratively investigate observational data from axSpA and psoriatic arthritis patients throughout Europe [17]. For the present study, we used data collected from 13 European registries (see Additional file 1) on axSpA patients diagnosed at an age of 18 years or older who initiated biological disease-modifying anti-rheumatic therapy with a tumour necrosis factor inhibitor (TNFi) until December 31, 2018. The data set provided information on the patient's first (and, possibly, further) TNFi treatment course and contained measurements of BASDAI, BASFI, and ASDAS-CRP components at up to three time points after the start of each TNFi (6-, 12-, and 24-months post-baseline) (see [18] for a prior study based on these data). For each composite score, we used all observations with complete component information regardless of treatment course or timing as the underlying distribution of interest. Table 1 provides an overview of the composite scores assessed.

Estimand and CC approach

The estimand, that is the quantity of interest whose estimation was to be assessed and compared between the different methods, was the expected value of the underlying composite score (BASDAI, BASFI, and ASDAS-CRP) distributions post-baseline. The estimand was estimated by the sample arithmetic mean.

The CC analysis was the reference method. A CC analysis provides unbiased estimates if the missingness is independent of the outcome variable (here the composite score) given the variables the estimation is conditioned on [10, 15]. With missingness completely at random (as chosen here), missingness and outcome variable are even unconditionally independent. It essentially means that a random sample of reduced size is obtained. The loss in precision of the arithmetic mean estimator with CC analysis compared to the original full sample is equal to the proportion of observations with missing component information (for a derivation see Additional file 1).

Imputation methods for missing component values

Imputation by the arithmetic mean of the remaining observed components for the observation (IMI), imputation by the arithmetic mean of all values observed in the sample for the component missing information (OMI), modified composite score formulae not using the components missing information (MF), and MI by full conditional specification [19] were all investigated and compared to the CC approach with respect to

Table 2 Imputation methods for missing component information

Method	Description
Individual mean imputation (IMI)	A missing component value is imputed by the arithmetic mean of the component values observed for that observation Example: A missing $Q_{BASDAI}1$ value is set to the arithmetic mean of the values observed for $Q_{BASDAI}2$ to $Q_{BASDAI}6$ resulting in a composite score of 0.2*(mean($Q_{BASDAI}2$, …, $Q_{BASDAI}6$) + $Q_{BASDAI}2$ + $Q_{BASDAI}3$ + $Q_{BASDAI}4$) + 0.1*($Q_{BASDAI}5$ + $Q_{BASDAI}6$)
Modified formula (MF)	The composite score is calculated as the weighted sum of the component values observed for the observation, where the weights are a scaled version of the components' original weight with scaling factor equal to the total sum of the components' original weights / the sum of original weights of observed components Example: In case of a missing $Q_{BASDAI}1$ value the composite score is derived as the weighted average of the values observed for $Q_{BASDAI}2$ to $Q_{BASDAI}6$ with weights of 0.25 for $Q_{BASDAI}2$ to $Q_{BASDAI}4$ and 0.125 for $Q_{BASDAI}5$ and $Q_{BASDAI}6$ resulting in a composite score of 0.25*($Q_{BASDAI}2$ + $Q_{BASDAI}3$ + $Q_{BASDAI}4$) + 0.125*($Q_{BASDAI}5$ + $Q_{BASDAI}6$)
Overall mean imputation (OMI)	A missing component value is imputed by the arithmetic mean of the values observed for that component in the other observations Example: A missing $Q_{BASDAI}1$ value is set to the arithmetic mean of all observed $Q_{BASDAI}1$ values in the sample resulting in a composite score of 0.2*(mean(observed $Q_{BASDAI}1$) + $Q_{BASDAI}2$ + $Q_{BASDAI}3$ + $Q_{BASDAI}4$) + 0.1*($Q_{BASDAI}5$ + $Q_{BASDAI}6$)
Multiple imputation (MI) of monotone missingness patterns using conditional specification	A comprehensive description of the method is provided in [10, 19] and guidance on its implementation with the R package mice is provided by [19, 20]. The type of imputation for a component was set to predictive mean matching and the imputation model consisted of all components fully observed for the first component imputed, all components fully observed plus the first component imputed for the second component imputed, and so on. Since all observations with missing component information in a particular sample missed information on the same components, the procedure simply imputed the components according to their numbering in the composite score formulae. We set the number of times we imputed equal to the sample's percentage of observations with missing component information [14], which ranged between 10 and 90%. For a sample with 90% of observations missing component information, we thus generated 90 completed data sets for which the missing component values were imputed as described above. Following calculation of individual composite scores, each of the 90 data sets provided an estimate of the composite score's expected value, which were then combined as described by [10, 19] to produce the MI estimate for the sample

performance. Following imputation of the missing component values, the composite scores were calculated. Table 2 provides an overview of the methods assessed and details of the implementation. ASDAS-CRP includes a CRP-based component equal to the log transformation of max (CRP, 2) + 1. For our assessments, we worked with this transformed CRP. For example, with OMI and a missing CRP measurement leading to a missing value for ln (max (CRP, 2) + 1), that missing value was imputed as the arithmetic mean of all ln (max (CRP, 2) + 1) values observed in the sample. Since ASDAS-CRP is not an average of components of equal scale and includes different sources (patient and laboratory), IMI and the MF method were not considered for this outcome.

Analytical assessments

IMI and the MF method are applied on the level of the individual observation. By setting the values of a particular set of components to missing for all observations in the EuroSpA RCN data and then applying the IMI or MF method to all of them allowed us to derive the underlying distribution of the score pertaining to each method and set of components. For BASDAI, which consists of six components, there were 62 different combinations or patterns of missing components with at least one observed. For BASFI with ten components there were 1022. For both methods, we characterised the 62 and 1022 score distributions by their arithmetic mean and variance and compared these with the arithmetic mean and variance of the true composite score distribution.

For each composite score, we then assessed three concrete missingness patterns (or scenarios) in more detail. Our choice of missingness patterns was guided by the assumption that the differences in the population means of components are the main determinant of the methods' performance. We therefore selected three patterns missing one component each, either the component with the smallest, the largest, or a population mean in the middle as observed in our underlying distributions. We investigated several settings per scenario and composite score that varied with respect to sample size and the proportion of observations missing information. The observations, for which component information was subsequently set to missing, were selected at random. The settings investigated differed for the different performance measures. The performance measures used were bias, variance, percent precision gain, and coverage of the two-sided 95% Wald-type confidence interval (CI).

Since bias and percent precision gain are invariant to changes in the sample size, we only varied the proportion of observations missing information. Variance and coverage, however, also depend on the sample size. We therefore assessed all combinations of sample size (200, 500, 1000, and 10,000) and proportion of observations missing information (0.05, 0.1, 0.2, 0.5, and 0.95). For more information on the performance measures see the later section on performance assessment.

Simulation

For OMI and MI, we utilized simulations to estimate the methods' performance in absence of analytical tractability.

Basic set-up

Each simulation was performed by repeatedly drawing with replacement from the underlying distributions of BASDAI, BASFI, or ASDAS-CRP to generate 2000 equally sized samples of independent and identically distributed data.

Introducing missing component information

For each sample, missing component information was introduced by selecting at random the desired number of observations for which certain component values were then set to missing. Given these conditions, both OMI and MI were expected to be unbiased and gain in precision was our main focus in terms of performance. Our choice of missingness patterns was guided as follows: The larger the extent of missingness is or the smaller the amount of information that is saveable, the less saving information should pay off in terms of precision. The extent of missingness is determined by the number of components missing and the weight they have in the composite score formula. We selected three missingness patterns or scenarios per composite score such that the extent of missingness increased (Table 3). For BASDAI and ASDAS-CRP two of the three patterns lacked information on a single component, either the component with the smallest or the largest weight. The third pattern missed information on two out of three or three out of five components, respectively. For BASDAI all but the two components in their respective weight groups with the smallest proportion of missing values in the EuroSpA RCN data were set missing. For ASDAS-CRP we selected the three components with the largest weight to be missing. For BASFI, with equal component weights, the components for missing information (either one, three, or six out of ten) were selected according to the observed proportion of missing values in the EuroSpA RCN data. For example, for BASDAI and scenario 3 with a sample size of 200 and a proportion of observations with missing component information of 50%, a sample consisted of 100 fully observed observations and 100 observations that missed information for the same four BASDAI components. For each composite score and given scenario, sample size, and proportion of observations with missing information we performed a separate simulation.

Simulation parameters

For each simulated scenario, we varied the sample size and the proportion of observations with missing information (ensuring that the number of completely observed observations was ≥ 100) [21]. For a sample size of 200, we selected 10%, 25%, 37.5%, and 50% of observations with missing component information. For 500, the proportions of observations with missing component information were 10%, 20%, 50%, and 80% and for 1000 10%, 50%, 75%, and 90%. In total, we investigated twelve parameter settings for each scenario and composite score, leading to 108 separate simulations.

Performance assessment

We assessed the performance of the methods by means of bias, variance (empirical and model-based), and

Table 3 Simulated missingness scenarios

	BASDAI		BASFI		ASDAS-CRP	
	Components with missing information[a]	Saveable amount (%)[b]	Components with missing information[a]	Saveable amount (%)[b]	Components with missing information[a]	Saveable amount (%)[b]
Scenario 1	$Q_{BASDAI}5$	90	$Q_{BASFI}10$	90	$Q_{BASDAI}6$	93.8
Scenario 2	$Q_{BASDAI}1$	80	$Q_{BASFI}10$, $Q_{BASFI}9$, $Q_{BASFI}8$	70	CRP	38.5
Scenario 3	$Q_{BASDAI}1$, $Q_{BASDAI}3$, $Q_{BASDAI}4$, $Q_{BASDAI}5$	30	$Q_{BASFI}10$, $Q_{BASFI}9$, $Q_{BASFI}8$, $Q_{BASFI}1$, $Q_{BASFI}3$, $Q_{BASFI}5$	40	CRP, $Q_{BASDAI}2$, patient global	13.9

[a] All other components were fully observed

[b] Of information for observations with missing component values. Calculated as the sum of weights of fully observed components/the sum of all weights. In case of ASDAS-CRP the sum of all weights is not equal to one

coverage of their estimator as well as some therefrom derived measures such as the percent relative error of the model-based versus the empirical variance and, for comparisons of methods, the percent precision gain of methods compared to the comparator method CC. For an overview and explanation of the performance measures used we refer to the Additional file 1 as well as [16] for further details. In case the performance measures are not analytically tractable or with difficulties only, simulations allow us to estimate them. We considered the chosen number of simulation repetitions to provide a suitable precision in estimating the performance measures for OMI and MI. To accommodate the uncertainty left, we reported the performance measures as estimates and two-sided 95% Wald-type CIs.

Software

All analyses were performed in R [22], version 4.2.0, with the RStudio IDE [23], server version 2022.02.2 + 485. For MI we used the package mice [20], version 3.15.0. To enable parallel computing, we made use of packages doParallel [24], version 1.0.17, and foreach [25], version 1.5.2. Figures were mainly produced using the ggplot2 package [26], version 3.4.0.

Results

Description of data

A total of 39,920 BASDAI, 20,680 BASFI, and 32,722 ASDAS-CRP observations with complete component information were available. Density plots and means for the composite scores are shown in Fig. 1 (see Figure S1, Additional file 2, for respective densities split by registry and Table S2, Additional file 2, for further information on the available data). The distributions for the different post-baseline timepoints were similar, which justified ignoring the post-baseline timing in our assessment from a clinical perspective. Figure 2 shows component means and variances (see Figures S2 and S3, Additional file 2, for overall and registry-split component densities). The differences between components in terms of means ranged over one to two units.

The number of observations with at least partial component information in the original EuroSpA RCN data varied for the different composite scores: BASFI components were assessed about half as often than BASDAI components, which were themselves assessed less often than ASDAS-CRP components. When obtained, however, BASFI had the lowest proportion of observations with missing component information, 1.6%, followed by BASDAI with 5.3% and ASDAS-CRP with 39.3% (see Tables S2/S3, Additional file 2). For ASDAS-CRP, which combines three distinct assessments (patient global assessment of disease, CRP, and BASDAI components), assuming the availability of at least one component means that its collection was intended, is questionable and very likely resulted in a comparatively large proportion of observations with missing component information. For all three composite scores, the missingness of components did not seem to be independent, i.e., missingness of several components per observation was more probable than expected under independent missingness

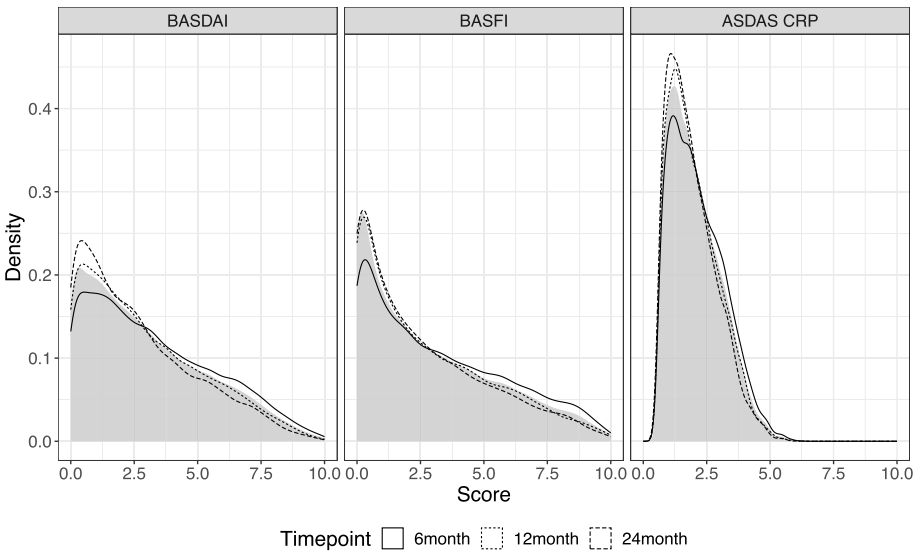

Fig. 1 Post-baseline composite score densities overall and by post-baseline timepoint. BASDAI: $n = 39{,}920$, mean = 3.02, BASFI: $n = 20{,}680$, mean = 2.9, ASDAS-CRP: $n = 32{,}722$, mean = 2.05. Shaded in grey the overall density. Densities were estimated with the function geom_density with default arguments from package ggplot2 [26]

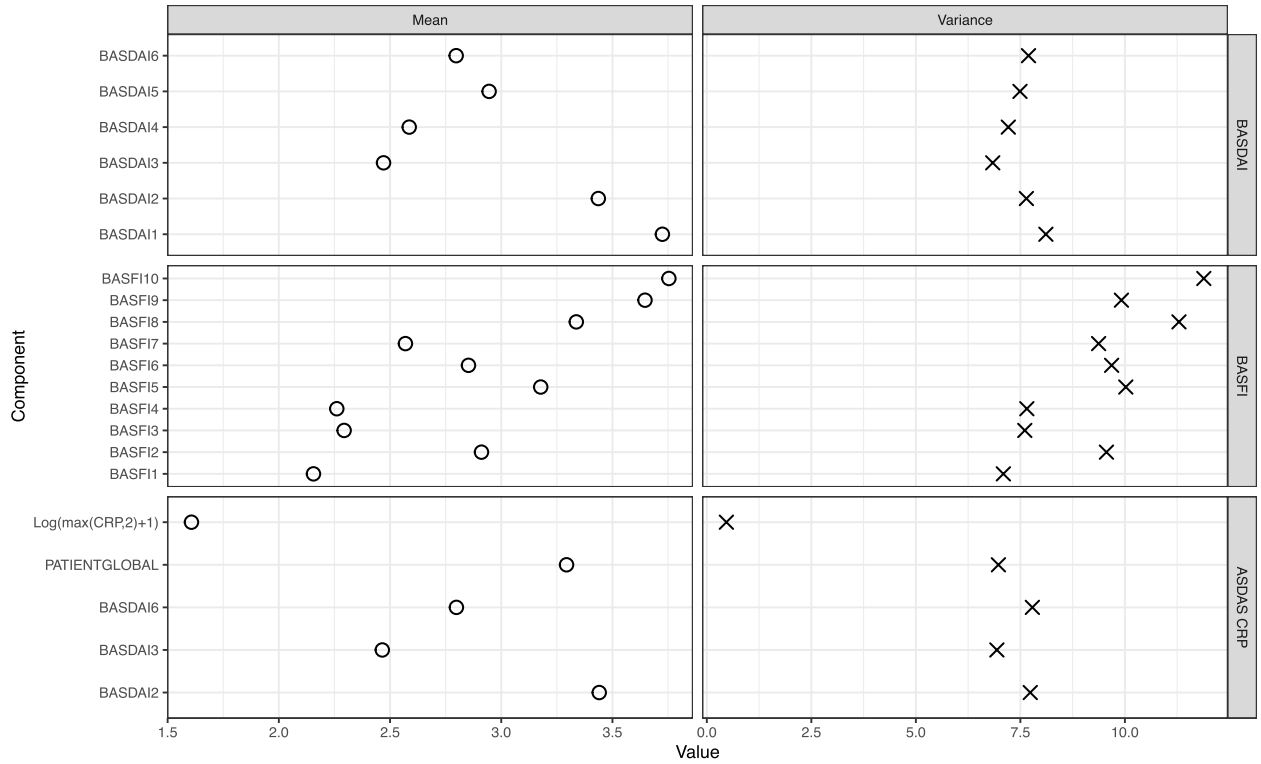

Fig. 2 Post-baseline component means and variances. BASDAI: $n = 39{,}920$, BASFI: $n = 20{,}680$, ASDAS-CRP: $n = 32{,}722$

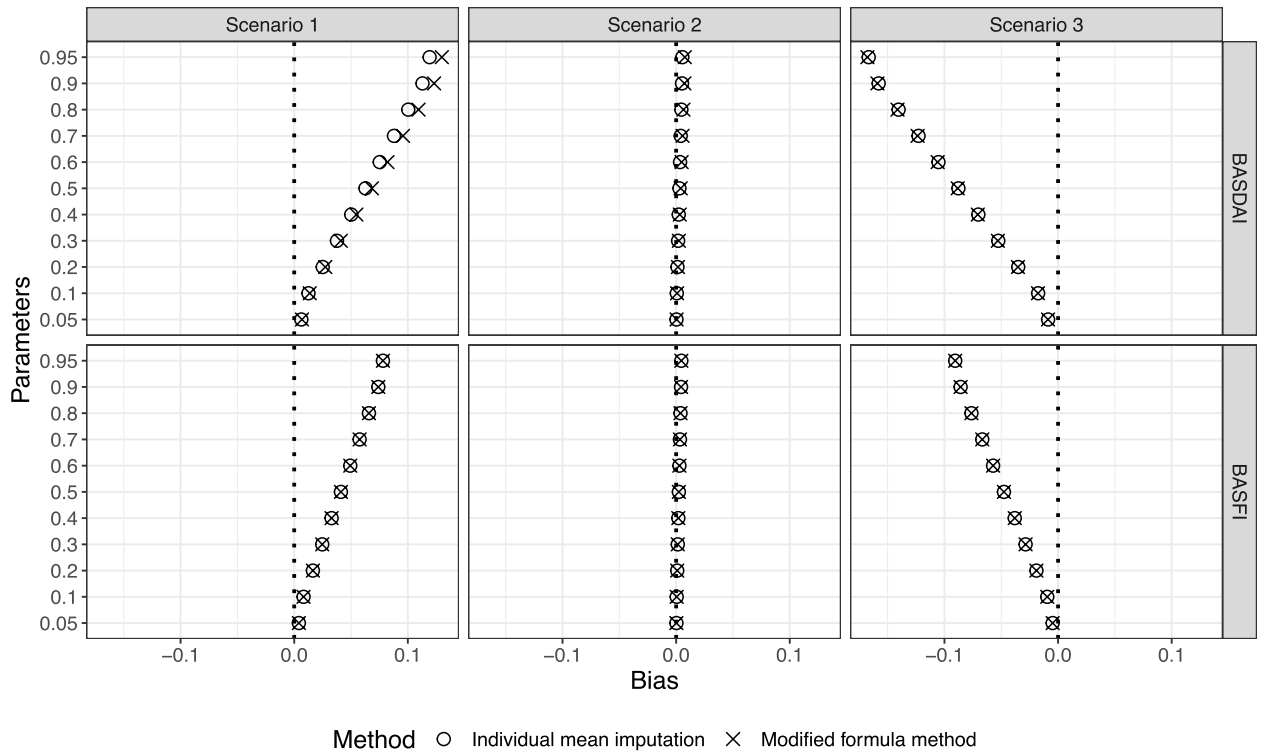

Fig. 3 Bias for IMI and MF. By composite score, scenario, and proportion of observations missing information. Population means (rounded): BASDAI: 3.02, BASFI: 2.9. For BASFI, IMI and MF are equivalent. Missing components according to scenario were: 1: $Q_{BASDAI}3$ or $Q_{BASFI}1$; 2: $Q_{BASDAI}5$ or $Q_{BASFI}6$; 3: $Q_{BASDAI}1$ or $Q_{BASFI}10$, respectively. IMI: individual mean imputation, MF: modified formula method

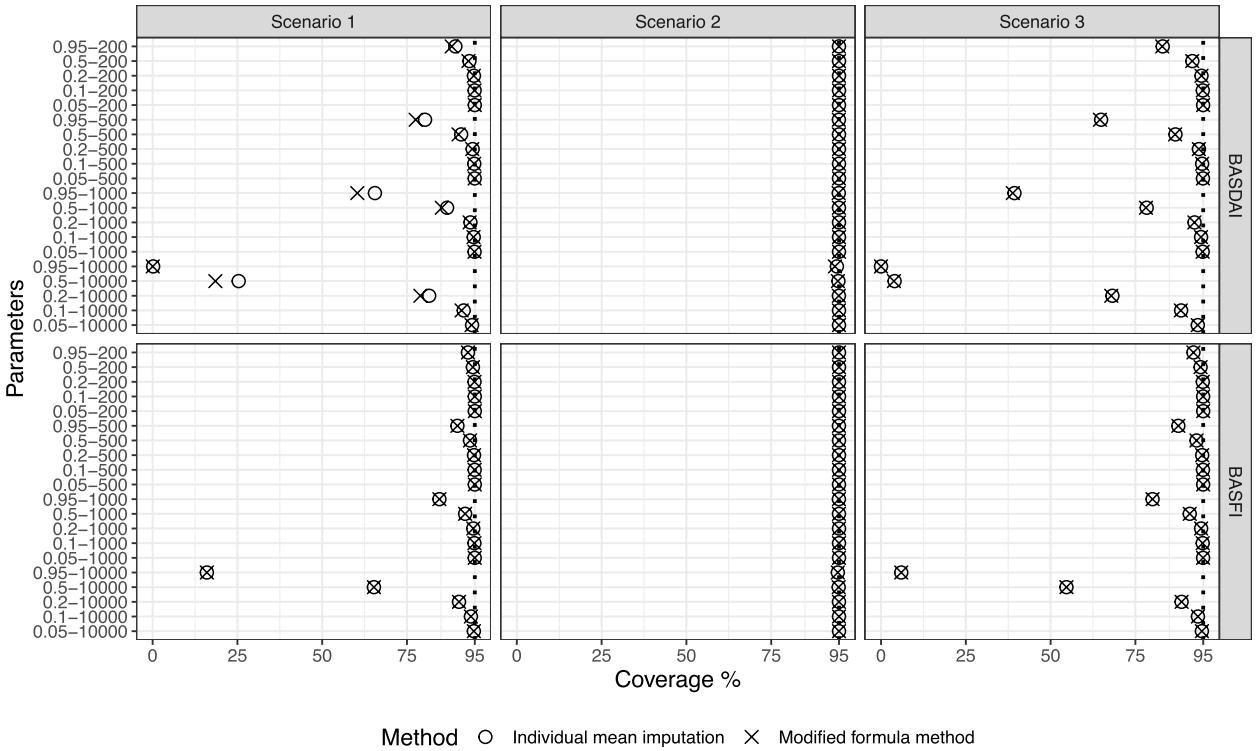

Fig. 4 Coverage of two-sided 95% Wald-type confidence interval for IMI and MF. By composite score, scenario, and setting. Population means (rounded): BASDAI: 3.02, BASFI: 2.9. All coverages were below 95% (dotted vertical reference line), albeit to a very small extent for some settings. For parameters, the first number refers to the proportion of observations missing information and the second number to the sample size. For BASFI, IMI and MF are equivalent. Missing components according to scenario were: 1: $Q_{BASDAI}3$ or $Q_{BASFI}1$; 2: $Q_{BASDAI}5$ or $Q_{BASFI}6$; 3: $Q_{BASDAI}1$ or $Q_{BASFI}10$, respectively. IMI: individual mean imputation, MF: modified formula method

of components. For BASDAI, for example, the majority of the partially observed observations lacked both component $Q_{BASDAI}1$ and $Q_{BASDAI}4$.

Analytical results

Individual mean imputation is just another modified formula method, which in the case of equal component weights as for BASFI is equivalent to our MF method (see Additional file 2 for a derivation). The distribution of modified composite scores should by default be assumed to be different from the actual composite score distribution unless the component distributions are the same. For our data, component distributions differ (Fig. 2 and Figures S2 and S3, Additional file 2) and so do the modified composite score distributions. Figures S4 and S5 from Additional file 2 show the expected value and variance of the post-baseline underlying distributions for both methods and all possible missingness patterns by composite score. The extent of the difference (less than one unit in absolute terms) between the modified composite score expected value and the BASDAI/BASFI expected value depends both on the identity of the missing components and the number of components missing, with a tendency for larger differences with more components missing. We also see that the variances differ. For each composite score, we selected three missingness patterns missing one component each, which covered the range of observed component expected values, for a more detailed assessment (Fig. 2 and Figure S4, Additional file 2). The missing components selected were either $Q_{BASDAI}3$ (scenario 1), $Q_{BASDAI}5$ (scenario 2), or $Q_{BASDAI}1$ (scenario 3) for BASDAI and $Q_{BASFI}1$ (scenario 1), $Q_{BASFI}6$ (scenario 2), or $Q_{BASFI}10$ (scenario 3) for BASFI. Bias (Fig. 3) and coverage (Fig. 4) in these cases turned out worse indeed if the component missing information showed the smallest or largest population mean and worsened with an increasing proportion of observations missing information or, in case of coverage, sample size. When the component missing information had a medium mean, bias was very small and coverage close to the nominal value over the range of investigated settings. For scenarios 1 and 3 and a sample size of 10,000, the coverage generally became unsatisfactory (closer to 90% than 95%) with 10% of observations missing information. With respect to precision, IMI and MF led to a gain compared to CC in all investigated settings (Fig. 5). The percent gain in

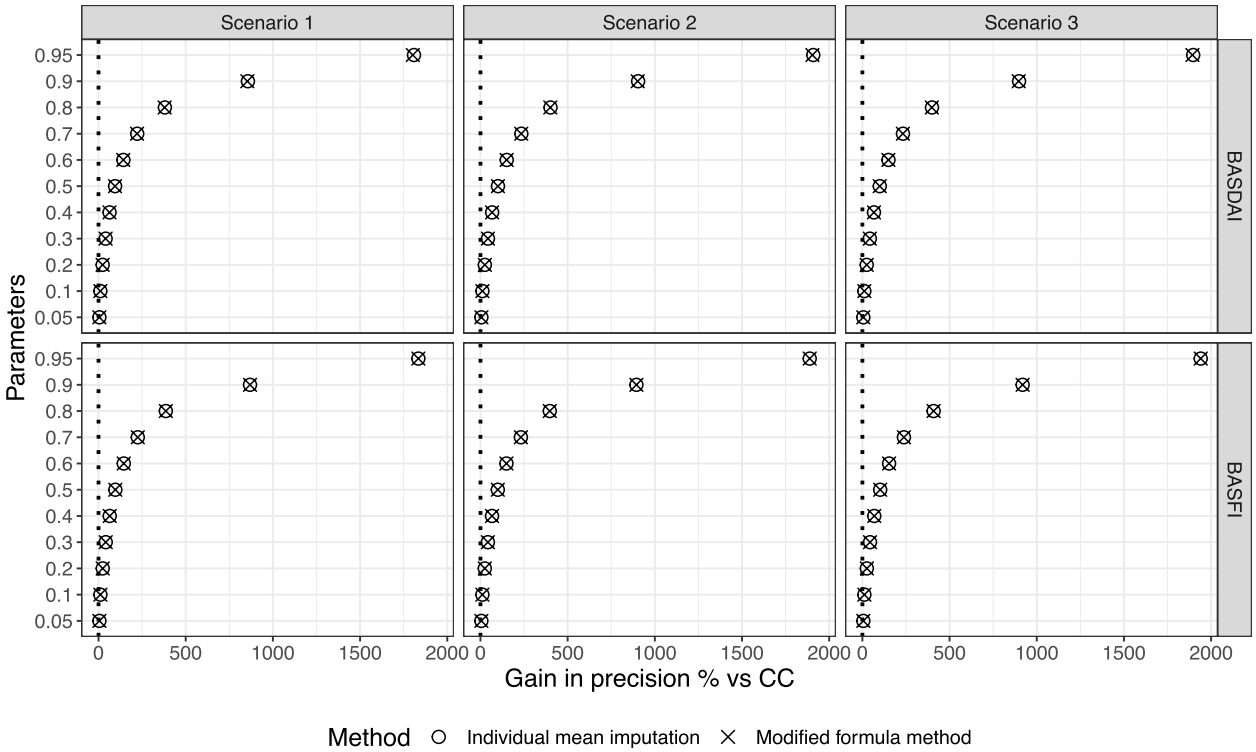

Fig. 5 Percent gain in precision compared to CC for IMI and MF. By composite score, scenario, and proportion of observations missing information. For BASFI, IMI and MF are equivalent. Missing components according to scenario were: 1: $Q_{BASDAI}3$ or $Q_{BASFI}1$; 2: $Q_{BASDAI}5$ or $Q_{BASFI}6$; 3: $Q_{BASDAI}1$ or $Q_{BASFI}10$, respectively. IMI: individual mean imputation, MF: modified formula method

precision was relatively small and of about the same magnitude for smaller proportions of observations missing information (5% and 10%) and increased disproportionately with larger proportions. The main determinant of the estimator's variance was the sample size, with little change over proportions of observations with missing information (see Figure S6, Additional file 2). There was little difference between composite scores with respect to the performance behaviour of IMI and MF over the range of settings examined.

Simulation results

All simulation repetitions provided estimates. Since our estimator was a simple sample arithmetic mean, we expected an approximate normal distribution even for the smallest sample size and largest amount of missingness, which was supported by the respective plots shown in Figure S7, Additional file 2. Figures S8 to S12, Additional file 2, display our estimates data.

The arithmetic mean of a sample of composite score values, obtained by imputing the missing component values by the OMI method, corresponds to the value obtained when applying the composite score formula to the sample means of components (for a derivation see Additional file 2). Since the expectation of a sample arithmetic mean is the same as the expectation of the variable averaged, OMI is unbiased. This is supported by the results of our simulation with respect to bias (Fig. 6). Only eight out of the 108 CIs for the bias did not contain zero, which is in line with a probability of non-coverage of 5% given unbiasedness. The empirical variance of the estimate obtained by OMI was remarkably smaller than the variance of the CC estimate with larger proportions of observations missing information (Fig. 7). However, since all missing values of a component are imputed by the same value, the method creates dependencies among the composite score values. Therefore, analysing them by standard means for a sample of independent observations is expected to result in variance underestimation and thus an inflated precision of the estimate. Variance underestimation was apparent as shown in Fig. 8 with percent relative error. As a result of the underestimation of the estimate's variance by the model (the estimated variance of a simple sample mean), OMI resulted in undercoverage to a possibly very large extent (Fig. 9). The extent of variance underestimation and undercoverage by OMI increased with the proportion of observations missing information, the component weight, as well as the number of components missing. However, for all but one of the 15 cases with only one missing component and a

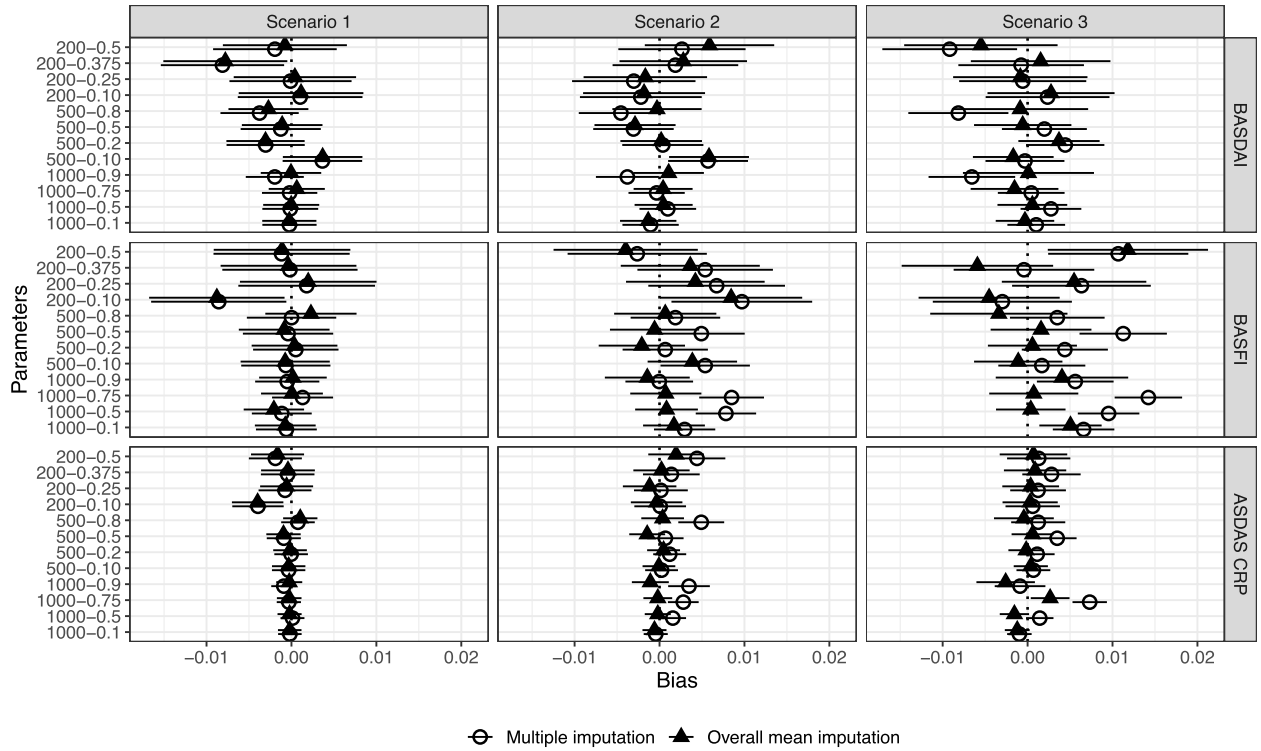

Fig. 6 Bias for MI and OMI. Estimates and two-sided 95% Wald-type CIs (simulation uncertainty) by composite score, scenario, and setting as obtained from simulations. For parameters, the first number refers to the sample size and the second number to the proportion of observations missing information. Missing components according to scenario were: 1: $Q_{BASDAI}5$ (BASDAI), $Q_{BASFI}10$ (BASFI), $Q_{BASDAI}6$ (ASDAS-CRP); 2: $Q_{BASDAI}1$ (BASDAI), $Q_{BASFI}8$, $Q_{BASFI}9$, and $Q_{BASFI}10$ (BASFI), CRP (ASDAS-CRP); 3: $Q_{BASDAI}1$, $Q_{BASDAI}3$, $Q_{BASDAI}4$, and $Q_{BASDAI}5$ (BASDAI), $Q_{BASFI}1$, $Q_{BASFI}3$, $Q_{BASFI}5$, $Q_{BASFI}8$, $Q_{BASFI}9$, and $Q_{BASFI}10$ (BASFI), CRP, $Q_{BASDAI}2$, and patient global (ASDAS-CRP). MI: multiple imputation, OMI: overall mean imputation

proportion of observations with missing information of 10%, the CI for coverage included 95% and undercoverage suggested to be small at most. In terms of precision, gain varied between 4.7% and 17.1% in these instances.

As expected, MI led to a remarkably more precise estimation compared to CC (Fig. 7 and Fig. 10). Its gain in precision increased with larger proportions of observations missing information, but to a lesser extent with increased numbers of components missing information. Similarly to OMI, gain varied between 4.6% and 17.2% in situations with one component missing information and a proportion of missing observations of 10%. In general, MI's empirical variance was smaller and its estimated gain in precision larger than for OMI. In terms of bias, we expected unbiasedness due to missingness completely at random. This was mostly supported by our simulation results (Fig. 6), though there may be some doubt in case of larger proportions of missing observations and more components missing information. Overall, 24 out of the 108 CIs for the bias did not contain zero, which is implausible with unbiasedness in all settings. The model-based variance estimation for the MI estimate was, however, in line with its empirical variance (Fig. 8), meaning that using MI in an actual case, where the variance of the estimate is estimated by the model, results in a proper variance estimate. Again, the performance may be questionable with large proportions of observations missing information. Coverage of MI (Fig. 9) reflected its performance with respect to bias and model-based variance estimation and its CIs included 95% in all but seven out of the 108 cases. Slight performance issues in case of more extensive missingness put aside, MI worked well.

Discussion

The composite scores assessed in this study and their specific derivations were developed to account for the multifaceted nature of the variable of interest, disease activity (BASDAI, ASDAS-CRP) [1, 3, 4] or functional impairment (BASFI) [2] in axSpA, and to allow comparability between clinical studies assessing these. Thus, it seems unreasonable to abandon them in the face of missing information. This, however, is exactly what is done with IMI and the MF method assessed in this study. These methods ignore components with missing information

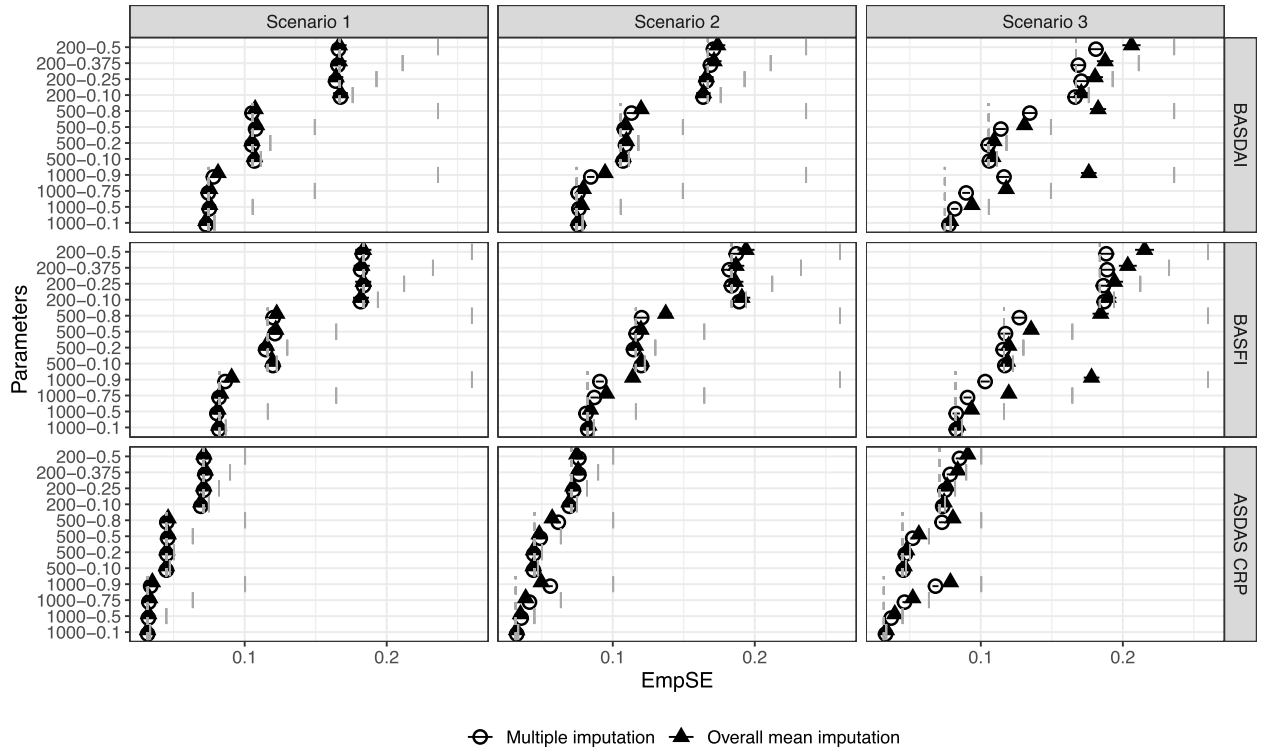

Fig. 7 Square root of empirical variance for MI and OMI. Estimates and two-sided 95% Wald-type CIs (simulation uncertainty) by composite score, scenario, and setting as obtained from simulations. Dashed reference lines show the variability obtainable without missing component information, whereas the solid reference lines show the variability obtained with the CC approach. For parameters, the first number refers to the sample size and the second number to the proportion of observations missing information. Missing components according to scenario were: 1: $Q_{BASDAI}5$ (BASDAI), $Q_{BASFI}10$ (BASFI), $Q_{BASDAI}6$ (ASDAS-CRP); 2: $Q_{BASDAI}1$ (BASDAI), $Q_{BASFI}8$, $Q_{BASFI}9$, and $Q_{BASFI}10$ (BASFI), CRP (ASDAS-CRP); 3: $Q_{BASDAI}1$, $Q_{BASDAI}3$, $Q_{BASDAI}4$, and $Q_{BASDAI}5$ (BASDAI), $Q_{BASFI}1$, $Q_{BASFI}3$, $Q_{BASFI}5$, $Q_{BASFI}8$, $Q_{BASFI}9$, and $Q_{BASFI}10$ (BASFI), CRP, $Q_{BASDAI}2$, and patient global (ASDAS-CRP). MI: multiple imputation, OMI: overall mean imputation

and calculate a modified score based on a reduced set of components.

In an actual study with missing component information, the data consists of a mixture of fully and partially observed observations. Applying IMI or the MF method to such a data set leads to a mixture of the actual composite score of interest (obtained for observations fully observed) and different modified scores (depending on the observed missingness patterns) and, thus, to a mixture of different distributions. The resulting mixture distribution must be considered different in expectation and variance from the distribution of the actual composite score, which was supported by our data. Even though the sample arithmetic mean in such a setting is an efficient (in the sense that no observations are lost) and unbiased (in case the requirements for unbiasedness of CC hold) estimate of its expected value, it cannot be assumed to be unbiased for the expected value of the composite score of interest. Any difference in expectation, however, will lead to substantial undercoverage and thus false conclusions with a sufficiently large sample size. The missingness patterns we assessed in more detail lacked information for one component only, yet depending on which component this was, bias and undercoverage varied markedly and could attain surprisingly large values. For a sample size of 10,000, as is easily achieved with collaborative studies, the coverage started to be unsatisfactory with 10% of observations missing information (Fig. 4). In light of the undercoverage and the potential for false conclusions it brings about, a gain in precision compared to CC analysis seems rather irrelevant. Furthermore, different studies will likely have a different mixture of the composite score and various modified scores and the distribution they target will differ even if applying the same method. This adds to observed differences across studies, although they all seem to assess the same composite score. We expect that modified formulae different from those assessed here suffer from the same problems. We suggest that IMI and the MF method should be viewed as failed attempts at saving sample size because they violate the composite score's conceptual basis and trade unbiasedness. Instead, we believe that our goal should be to save the information we gathered with respect to the composite score of interest and stay faithful to it. On

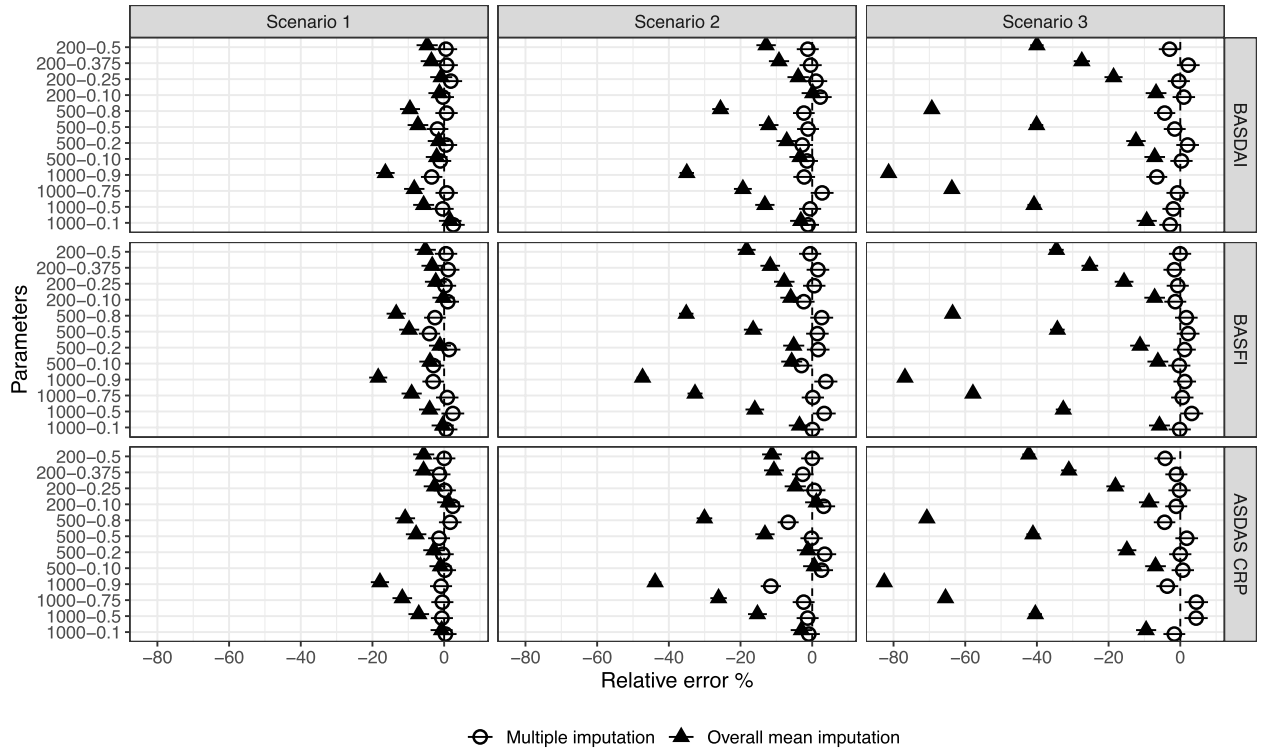

Fig. 8 Percent relative error of model-based variance estimation for MI and OMI. Estimates and two-sided 95% Wald-type CIs (simulation uncertainty) by composite score, scenario, and setting as obtained from simulations. For parameters, the first number refers to the sample size and the second number to the proportion of observations missing information. Missing components according to scenario were: 1: $Q_{BASDAI}5$ (BASDAI), $Q_{BASFI}10$ (BASFI), $Q_{BASDAI}6$ (ASDAS-CRP); 2: $Q_{BASDAI}1$ (BASDAI), $Q_{BASFI}8$, $Q_{BASFI}9$, and $Q_{BASFI}10$ (BASFI), CRP (ASDAS-CRP); 3: $Q_{BASDAI}1$, $Q_{BASDAI}3$, $Q_{BASDAI}4$, and $Q_{BASDAI}5$ (BASDAI), $Q_{BASFI}1$, $Q_{BASFI}3$, $Q_{BASFI}5$, $Q_{BASFI}8$, $Q_{BASFI}9$, and $Q_{BASFI}10$ (BASFI), CRP, $Q_{BASDAI}2$, and patient global (ASDAS-CRP). MI: multiple imputation, OMI: overall mean imputation

these grounds, we reject the IMI and the MF methods as valid approaches to assess composite scores like BASDAI and BASFI in research.

According to earlier work [7], it may be likely that researchers are not aware of the fact that IMI or the MF method are techniques to deal with missing information but rather regard them as a feature of the composite score definition. We presume that the decision to allow for some component missingness is not based on an assessment of its usefulness for research purposes but rather taken out of necessity to use a composite score for the treatment of individual patients. This view is supported by other work [27] that aimed to provide guidance on the maximum number of components with a missing value for which a valid assessment of an individual patient's disease activity in terms BASDAI or physical function in terms of BASFI can still be obtained. Given this context, we believe that it is crucial to be aware of and distinguish between the applications and tailor the requirements accordingly.

In our study, imputing a missing component value with its sample mean seemed to be a reasonable approach as it did not introduce bias with respect to the composite score's expected value and increased precision compared to a CC approach. However, OMI suffers from another important issue: dependence among observations. This is problematic because proper handling of the dependence in the analysis of a sample of composite scores is demanding and variance underestimation and undercoverage are therefore expected. Indeed, as the proportion of missing observations or the number of components missing information increased, we observed a remarkable extent of variance underestimation and undercoverage under the assumption of independence of observations (Fig. 9 and Fig. 10). Furthermore, OMI generally led to a smaller increase in precision compared to CC than obtainable with MI. The deficiencies of OMI have been highlighted in previous work [7, 9]. Nonetheless, we believe that our assessment provides additional insight regarding how and why it fails for composite scores.

MI saved the available information about the composite score, contained in the observed components of observations with partial information, and led to a more precise

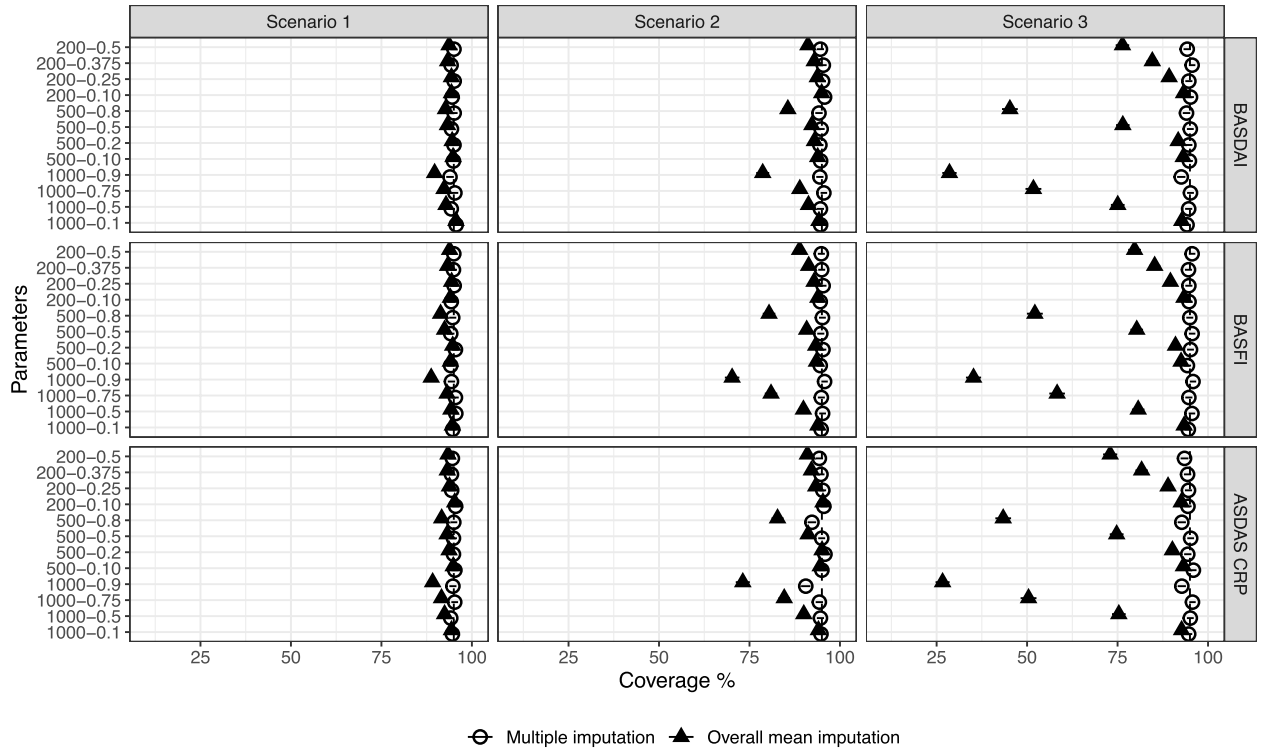

Fig. 9 Coverage of two-sided 95% Wald-type CI for MI and OMI. Estimates and two-sided 95% Wald-type CIs (simulation uncertainty) by composite score, scenario, and setting as obtained from simulations. For parameters, the first number refers to the sample size and the second number to the proportion of observations missing information. Missing components according to scenario were: 1: $Q_{BASDAI}5$ (BASDAI), $Q_{BASFI}10$ (BASFI), $Q_{BASDAI}6$ (ASDAS-CRP); 2: $Q_{BASDAI}1$ (BASDAI), $Q_{BASFI}8$, $Q_{BASFI}9$, and $Q_{BASFI}10$ (BASFI), CRP (ASDAS-CRP); 3: $Q_{BASDAI}1$, $Q_{BASDAI}3$, $Q_{BASDAI}4$, and $Q_{BASDAI}5$ (BASDAI), $Q_{BASFI}1$, $Q_{BASFI}3$, $Q_{BASFI}5$, $Q_{BASFI}8$, $Q_{BASFI}9$, and $Q_{BASFI}10$ (BASFI), CRP, $Q_{BASDAI}2$, and patient global (ASDAS-CRP). MI: multiple imputation, OMI: overall mean imputation

estimation compared to CC (Fig. 10). Moreover, it was faithful to the composite score formula of interest and, provided missingness was not considerable (in terms of proportion of observations missing information or number of components missing), resulted in an unbiased estimate (Fig. 6) with a sound sample variance estimation (Fig. 8). Interestingly, but not surprisingly, it was not just the number of components observed but also their relative weight that seemed to play a role for the precision gained with MI; the larger the relative weight of the observed components was, the more precision was gained.

We chose a simple imputation model for MI, consisting of just the composite score components in a predictive mean matching, to assess in a"pure" way what we can gain in terms of precision by just considering the available component information. In an actual study, additional information about the distribution of the component with missing values may be available from other variables that were observed. These should be included in the imputation model as they can further decrease the uncertainty about the missing component value and thereby increase MI's precision [10, 14, 15, 19, 28]. This is a feature unique to MI; none of the other methods assessed could benefit from the availability of such additional information. The uniqueness of this feature contributed to our decision to impute missing component values solely based on the information about them contained in other components.

CC was the comparator method and we aimed for another method that shares CC's unbiasedness but not its inefficiency. We consider MI at the component level to have been successful in doing so. However, there are two important aspects in CC's favour. First, there is a huge gap between CC and MI with respect to the effortlessness of implementation. According to [14], MI is a powerful technique that, however, must be used with understanding and care. MI involves multiple steps and entails many potential pitfalls [9, 14]. Compared to this, there is nothing easier and more time-saving than to throw out observations missing some of the required information. Second, no matter how large the proportion of observations lost due to missing component information, if unbiased, CC is based on a sample of independent observations representative of the underlying distribution and its estimate stays unbiased and its variance estimation sound leading to proper coverage.

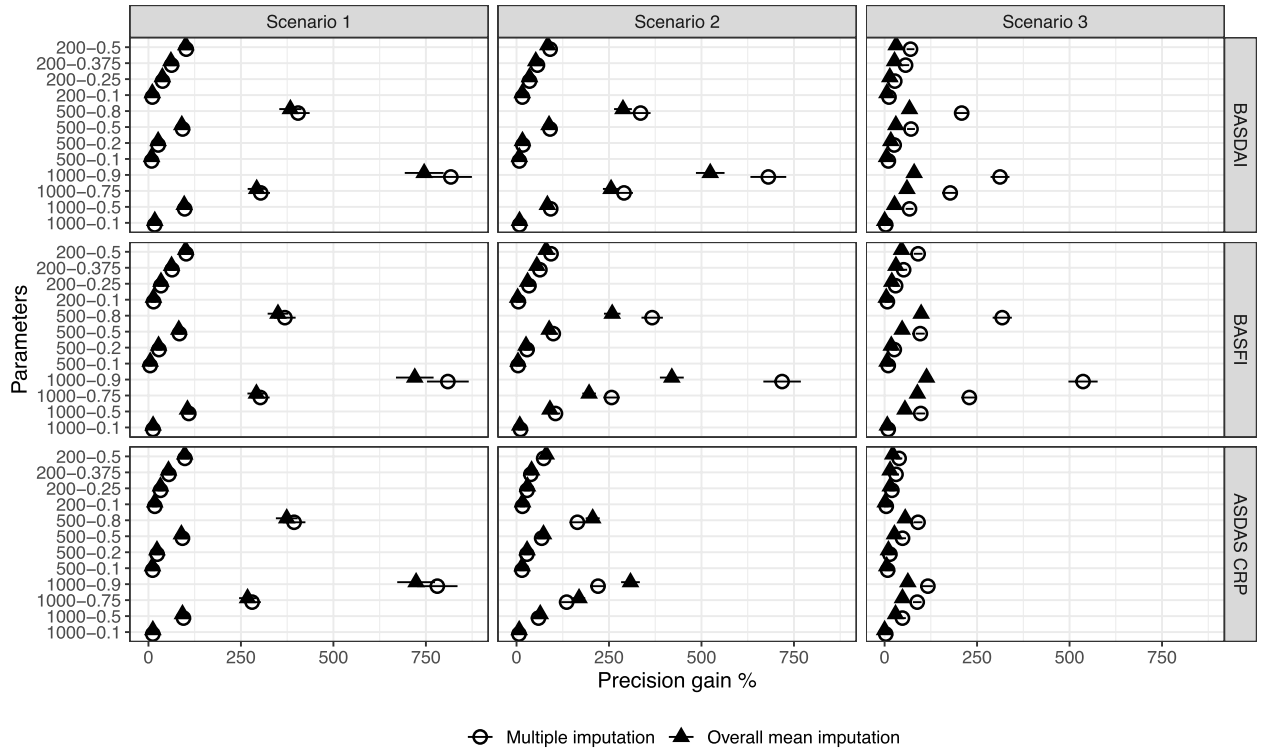

Fig. 10 Percent gain in precision compared to CC for MI and OMI. Estimates and two-sided 95% Wald-type CIs (simulation uncertainty) by composite score, scenario, and setting as obtained from simulations. For parameters, the first number refers to the sample size and the second number to the proportion of observations missing information. Missing components according to scenario were: 1: $Q_{BASDAI}5$ (BASDAI), $Q_{BASFI}10$ (BASFI), $Q_{BASDAI}6$ (ASDAS-CRP); 2: $Q_{BASDAI}1$ (BASDAI), $Q_{BASFI}8$, $Q_{BASFI}9$, and $Q_{BASFI}10$ (BASFI), CRP (ASDAS-CRP); 3: $Q_{BASDAI}1$, $Q_{BASDAI}3$, $Q_{BASDAI}4$, and $Q_{BASDAI}5$ (BASDAI), $Q_{BASFI}1$, $Q_{BASFI}3$, $Q_{BASFI}5$, $Q_{BASFI}8$, $Q_{BASFI}9$, and $Q_{BASFI}10$ (BASFI), CRP, $Q_{BASDAI}2$, and patient global (ASDAS-CRP). MI: multiple imputation, OMI: overall mean imputation

This holds, unless approximate methods were used that would not hold given the distribution and size of the sample (we also acknowledge that one may end up with a sample too small for a valid covariate-adjusted analysis). With increasing missingness in the data, MI, on the other hand, must be viewed cautiously [21, 28]. In our study, we started to note possible issues with coverage (Fig. 9) for proportions of observations missing information greater than 50%. As a result, researchers may thus just need to accept their faulty data and the low precision arising from a simple CC approach. Even with little missingness, the difficulties arising from implementing a proper MI approach may not outweigh the expected gain in precision and a CC approach could be an easy and reasonable option to choose.

Our findings for the different composite scores were remarkably similar. We ascribe this to the fact that they are all weighted sums of components. We ascribe differences between composite scores in the extent of performance issues primarily to features of their formula (number and relative weighing of components, sum of weights) and the investigated scenarios (for ASDAS-CRP we investigated scenarios with less information to be saved).

Finally, our investigations were based on a large data set combining composite score information of patients undergoing TNFi therapy from multiple European countries. We hypothesize that our conclusions are also valid for underlying distributions that differ in location or shape as observed for example between countries (Figure S1, Additional file 2).

Limitations

Estimating a marginal expected value may rarely be the primary goal and missingness completely at random seldomly fulfilled. Typically, we are interested in estimating model parameters and willing to assume missingness at random. It is important to note that the parameters of a linear regression correspond to the respective differences between conditional expected values of the outcome. In our view, the relative performance of methods in estimating the expected value of a composite score distribution should not depend on the conditioning of the targeted distribution and apply accordingly to differences between expected values. For example, as long as

both MI and CC analysis are unbiased, the fundamental conclusions drawn from our investigations should be extrapolatable to more realistic conditional situations as encountered in linear regression.

Conclusions

Of all the methods assessed in this study, we consider MI at the component level as the only method capable of retaining CC's unbiasedness and leading to increased precision in estimating the expected value of common composite scores (BASDAI, BASFI, and ASDAS-CRP) used in axSpA research. Unsurprisingly, the gain in precision with MI diminished with an increasing extent of component missingness at the level of observations.

MI is susceptible to incorrect implementation and its performance questionable with increasing missingness whereas CC is basically implementation error-free and its performance, apart from loss in precision, unwavering even with massive missingness. In the absence of suspected bias with both approaches in the face of missing component information for a composite score outcome, the gain in precision arising from an MI over a CC approach needs to be carefully weighed against MI's demanding implementation and susceptibility for impaired performance with more extensive missingness and CC may be chosen as the safer option.

Abbreviations

ASAS	Assessment of SpondyloArthritis international Society
ASDAS (-CRP)	Ankylosing Spondylitis Disease Activity Score (based on C-Reactive Protein)
axSpA	Axial SpondyloArthritis
BASDAI	Bath Ankylosing Spondylitis Disease Activity Index
BASFI	Bath Ankylosing Spondylitis Functional Index
CC	Complete Cases
CI	Confidence Interval
IMI	Individual Mean Imputation
MF	Modified Formula
MI	Multiple Imputation
OMI	Overall mean imputation
RCN	Research Collaboration Network

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12874-025-02515-3.

Additional file 1. Supplementary methods for Handling of missing component information for common composite score outcomes used in axial spondyloarthritis research when complete-cases analysis is unbiased. Description: Additional information with respect to methods like a list of participating registries, mathematical derivations, and detailed descriptions of performance measures.

Additional file 2. Supplementary results for Handling of missing component information for common composite score outcomes used in axial spondyloarthritis research when complete-cases analysis is unbiased. Description: Additional information with respect to results such as the underlying composite score data, performance measures, and simulated data.

Acknowledgements

We thank all the patients and health care professionals who collected the data for the individual registries participating in this study (see Additional file 1) as well as EuroSpA's coordinating centre who provided the necessary infrastructure and support for collecting, consolidating, and analysing the data.

Authors' contributions

The study protocol was drafted by SG, LMØ, DDG, MLH, AS, GJ, MR, and MN and revised and approved by all authors. Data analyses were done by CP and MR and the manuscript was drafted by MR, CP, SG, LMØ, DDG, AS, and MLH. All authors have contributed substantially to the acquisition of data, revised the manuscript and approved the final submitted version.

Funding

This work and the EuroSpA Research Collaboration Network were 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 (see Additional file 1) and made available for secondary use through the EuroSpA Research Collaboration Network. Relevant patient level data may be made available on reasonable request to the corresponding author, but will require approval from all contributing registries.

Declarations

Ethics approval and consent to participate

All patient data were collected in accordance with national legal and regulatory requirements in the different countries (including obtaining written informed consent from participating patients where required). The study was approved by the respective national Data Protection Agencies or Ethical Committees according to legal regulatory requirements in the participating countries.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

[1]Statistics Group, Swiss Clinical Quality Management (SCQM) Foundation, Aargauerstrasse 250, Zurich 8048, Switzerland. [2]Copenhagen Centre for Arthritis Research (COPECARE), Centre for Rheumatology and Spine Diseases, Centre for Head and Orthopaedics, Rigshospitalet, Glostrup, Denmark. [3]Department of Medicine Solna, Clinical Epidemiology Division, Karolinska Institutet, Stockholm, Sweden. [4]Copenhagen Centre for Arthritis Research (COPECARE) and the DANBIO Registry, Centre for Rheumatology and Spine Diseases, Centre for Head and Orthopaedics, Rigshospitalet, Glostrup, Denmark. [5]Department of Rheumatology, Geneva University Hospital, Geneva, Switzerland. [6]Aberdeen Centre for Arthritis and Musculoskeletal Health (Epidemiology Group), University of Aberdeen, Aberdeen, UK. [7]Department of Rheumatology, Aarhus University Hospital, Aarhus, Denmark. [8]Department of Clinical Sciences Lund, Rheumatology, Skåne University Hospital, Lund University, Lund, Sweden. [9]Institute of Rheumatology, Prague, Czech Republic. [10]Division of Rheumatology, Department of Internal Medicine, School of Medicine, Kocaeli University, Kocaeli, Turkey. [11]Sociedade Portuguesa de Reumatologia, Reuma.Pt, Lisbon, Portugal. [12]Department of Rheumatology, University Hospital Zurich, University of Zurich, Zurich, Switzerland. [13]Department Rheumatology and Immunology, Inselspital - University Hospital Bern, Bern, Switzerland. [14]Center for Treatment of Rheumatic and Musculoskeletal Diseases (REMEDY), Diakonhjemmet Hospital, Oslo, Norway. [15]Section for Rheumatology, Department for Neurology, Rheumatology and Physical Medicine, Helse Førde, Førde, Norway. [16]Department of Rheumatology, Kuopio University Hospital, Kuopio, Finland. [17]Rheumatology, Inflammation Center, Helsinki University Hospital and University of Helsink, Helsinki, Finland. [18]Faculty of Medicine, University of Ljubljana, Ljubljana, Slovenia. [19]Faculty of Medicine, University of Iceland, Reykjavik, Iceland. [20]Centre for Rheumatology Research, Landspitali University Hospital, Reykjavik, Iceland. [21]Rheumatology, Radboud University Medical

Centre, Nijmegen, The Netherlands. [22]Amsterdam UMC, Department of Rheumatology & Clinical Immunology and Department of Experimental Immunology, Amsterdam Institute for Infection & Immunity, University of Amsterdam, Amsterdam, The Netherlands.

Received: 28 November 2023 Accepted: 19 February 2025
Published online: 28 February 2025

References

1. 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.
2. 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(12):2281–2285.
3. 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. 2009:68:18–24.
4. van der Heijde D, Lie E, Kvien T, Sieper J, Van den Bosch F, Listing J, et al. ASDAS, a highly discriminatory ASAS-endorsed disease activity score in patients with ankylosing spondylitis. Ann Rheum Dis. 2009:68:1811–8.
5. Jones GT, Dean LE, Pathan E, Hollick RJ, Macfarlane GJ. Real-word evidence of TNF inhibition in axial spondyloarthritis: can we generalise the results from clinical trials? Ann Rheum Dis. 2020:79:914–9.
6. Marzo-Ortega H, Sieper J, Kivitz AJ, Blanco R, Cohen M, Pavelka K, et al. 5-year efficacy and safety of secukinumab in patients with ankylosing spondylitis: end-of-study results from the phase 3 MEASURE 2 trial. Lancet Rheumatol. 2020;2(6):e339–46.
7. Schafer JL, Graham JW. Missing data: our view of the state of the art. Psychol Methods. 2002;7(2):147.
8. Braun J, McHugh N, Singh A, Wajdula JS, Sato R. Improvement in patient-reported outcomes for patients with ankylosing spondylitis treated with etanercept 50 mg once-weekly and 25 mg twice-weekly. Rheumatology. 2007;46(6):999–1004.
9. Horton NJ, Kleinman KP. Much ado about nothing: A comparison of missing data methods and software to fit incomplete data regression models. Am Stat. 2007;61(1):79–90.
10. Carpenter JR, Kenward MG. Multiple imputation and its application. 1st ed. Hoboken: Wiley & Sons; 2013.
11. Rombach I, Gray AM, Jenkins C, Murray DW, Rivero-Arias O. Multiple imputation for patient reported outcome measures in randomized controlled trials: advantages and disadvantages of imputing at the item, subscale or composite score level. BMC Med Res Methodol. 2018;18:87.
12. O'Keeffe AG, Farewell DM, Tom BDM, Farewell VT. Multiple imputation of missing composite outcomes in longitudinal data. Stat Biosci. 2016;8:310–32.
13. Simons CL, Rivero-Arias O, Yu LM, Simon J. Multiple imputation to deal with missing EQ-5D-3L data: should we impute individual domains or the actual index? Qual Life Res. 2015;24(4):805–15.
14. White IR, Royston P, Wood AM. Multiple imputation using chained equations: issues and guidance for practice. Stat Med. 2011;30(4):377–99.
15. Hughes RA, Heron J, Sterne JAC, Tilling K. Accounting for missing data in statistical analyses: multiple imputation is not always the answer. Int J Epidemiol. 2019;48(4):1294–304.
16. Morris TP, White IR, Crowther MJ. Using simulation studies to evaluate statistical methods. Stat Med. 2019;38:2074–102.
17. The EuroSpA Research Collaboration Network. URL https://eurospa.eu (2016). Accessed 2 Feb 2023.
18. Ornbjerg LM, Linde L, Georgiadis S, Rasmussen SH, Lindström U, Askling J, et al. Predictors of ASDAS-CRP inactive disease in axial spondyloarthritis during treatment with TNF-inhibitors: Data from the EuroSpA collaboration. Semin Arthritis Rheum. 2022;56: 152081.
19. Van Buuren S. Flexible imputation of missing data. 2nd ed. London: Chapman and Hall/CRC; 2012.
20. van Buuren S, Groothuis-Oudshoorn K. mice: multivariate imputation by chained equations in R. J Stat Softw. 2011;45(3):1–67.
21. Kleinke K. Multiple imputation by predictive mean matching when sample size is small. Methodology. 2018;14:3–15.
22. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2022. URL https://www.r-project.org/.
23. RStudio Team. RStudio: Integrated development for R. RStudio, PBC, Boston, MA. 2020. URL http://www.rstudio.com.
24. Microsoft Corporation, Weston S. doParallel: foreach parallel adaptor for the 'parallel' package. 2022. URL https://cran.r-project.org/package=doParallel.
25. Microsoft Corporation, Weston S. foreach: provides foreach looping construct. 2022. URL https://cran.r-project.org/package=foreach.
26. Wickham H. ggplot2: elegant graphics for data analysis. 2nd ed. New York: Springer; 2016.
27. Ramiro S, van Tubergen A, van der Heijde D, van den Bosch F, Dougados M, Landewé R. How to deal with missing items in BASDAI and BASFI. Rheumatology. 2014;53(2):374–6.
28. 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. 2009;338: b2393.

Publisher's Note