

# Multiple Imputation of Missing Values Using the Response Function Method Based on a Data Set of the Health Assessment Questionnaire Disability Index

Sağlık Değerlendirme Anketi Özürülük İndeksi Veri Setinde Eksik Veriler Yerine Yanıt Fonksiyonu Yöntemi ile Çoklu Değer Ataması

Beyza DOĞANAY ERDOĞAN,<sup>1</sup> Atilla H. ELHAN,<sup>1</sup> Hakan DEMİRTAŞ,<sup>2</sup> Derya ÖZTUNA,<sup>1</sup>  
Ayşe A. KÜÇÜKDEVECİ,<sup>3</sup> Şehim KUTLAY<sup>3</sup>

<sup>1</sup>Department of Biostatistics, Medical Faculty of Ankara University, Ankara, Turkey;

<sup>2</sup>University of Illinois at Chicago, Division of Epidemiology and Biostatistics, Chicago, Illinois, USA;

<sup>3</sup>Department of Physical Medicine and Rehabilitation, Medical Faculty of Ankara University, Ankara, Turkey

**Objectives:** This study aims to investigate how imputing missing values in data obtained from the Health Assessment Questionnaire Disability Index (HAQ-DI) influences the bias and precision of patient disability measurements.

**Patients and methods:** Hypothetical missing data sets were created by deleting item responses completely at random from the original data set with three missingness proportions (0.10, 0.30 and 0.50). Multiple imputation was carried out using the response function method for each hypothetical data set containing the missing values. The Rasch model was used to estimate the patients' latent trait levels for the original data, the hypothetical incomplete data sets, and the multiple imputed data sets. Then the estimates from the hypothetical missing data sets and the multiple imputed data sets were compared with those of the original data set.

**Results:** A bias in disability estimates was observed, particularly as the missingness proportion increased for both the incomplete and imputed data; however, this bias was indiscernible even for the 0.50 proportion of missingness. In terms of the uncertainty of the disability estimates, the imputed data had a higher precision of estimates than the incomplete data.

**Conclusion:** When researchers encounter missingness in data collected with the HAQ-DI, the response function imputation could be a convenient approach to impute missing values in order to improve the precision of the patient disability level estimates.

Key words: Missing data analysis; multiple imputation; partial credit model; Rasch analysis; response function.

**Amaç:** Bu çalışmada, Sağlık Değerlendirme Anketi Özürülük İndeksi'nden (HAQ-DI) elde edilen verideki eksik değerler yerine değer atamanın hasta özürülük ölçümlerinin yanlılık ve duyarlılığını nasıl etkilediği araştırıldı.

**Hastalar ve yöntemler:** Hipotetik eksik veri setleri oluşturmak için orijinal veri setindeki madde yanıtları tamamen rastgele eksik olmak üzere, üç farklı oranda (0.10, 0.30 ve 0.50) silindi. Eksik veri içeren her hipotetik veri seti için yanıt fonksiyonu yöntemi ile çoklu değer ataması yapıldı. Tam veri, hipotetik olarak oluşturulmuş eksik veri setleri ve değer ataması yapılmış veri setleri için Rasch modeli kullanılarak, hasta özürülük düzeyleri kestirildi. Eksik veri setleri ve değer ataması yapılmış veri setlerinden bulunan kestirimler tam veriden bulunanlar ile kıyaslandı.

**Bulgular:** Hem eksik veri durumdan hem de değer ataması yapılmış durumdan bulunan özürülük düzeyi kestirimlerinde, özellikle eksik veri oranı arttıkça, bir miktar yanlılık gözlenmiş olsa da, bu yanlılık eksik veri oranı 0.50 olduğunda dahi kabul edilebilir düzeyde idi. Değer ataması yapılmış veriden bulunan kestirimlerin duyarlılığı, eksik değer içeren veriden bulunanlara göre daha yüksek bulundu.

**Sonuç:** Sağlık Değerlendirme Anketi Özürülük İndeksi ile toplanan veride eksik madde yanıtları bulunduğu, bu eksikler yerine yanıt fonksiyonu ile değer atama yapılması, hastaların özürülük düzeyi kestirimlerinin duyarlılığının artırılması için önerilebilir.

Anahtar sözcükler: Kayıp veri analizi, çoklu değer atama, kısmi kredi modeli, Rasch analizi; yanıt fonksiyonu.

**Received:** October 31, 2012 **Accepted:** November 5, 2012

**Correspondence:** Beyza Doğanay Erdoğan, PhD. Ankara Üniversitesi Tıp Fakültesi, Biyoistatistik Anabilim Dalı, 06100 Sıhhiye, Ankara, Turkey.

Tel: +90 312 - 595 81 39 e-mail: beyzadoganay@gmail.com

©2013 Turkish League Against Rheumatism. All rights reserved.

Musculoskeletal diseases are chronic and disabling disorders, and assessment of patient disability levels during the planning and monitoring stages of the therapeutic approaches is essential for outcome measurement. Scales are commonly used to assess the latent traits which cannot be measured directly, such as disability; hence, having valid and reliable outcome scales is crucial. Because many of these scales are patient-reported, it is highly possible to observe missing data in such questionnaires for the following reasons: (i) Skipping one or more items unintentionally, (ii) Leaving some items blank to answer later but then forgetting about them, (iii) Including items that are not applicable to the patient, (iv) Losing interest in the questionnaire, and (v) Having limited time to respond to the items so some questions are skipped. The missing data is defined as “missing completely at random (MCAR)” when the missingness is unrelated to either the missing responses themselves, the observed responses of other items, the observed covariates, or the values of variables which are not included in the study.

The approaches for handling missingness in scales vary, but deleting the items which have missing responses, using all the available data, and imputing the missing values are all possible. Actually, there are some problems with the first two approaches. For the first approach, if the item that has missing data is crucial for the assessment, deleting that item may cause a violation of the content validity of that scale. For the second approach, because of the reduced sample size, the reliability of the patients’ estimated measurements (i.e. disability) will be underestimated. Therefore, the third approach, imputing the missing values by using the available data, is more appealing than the other approaches in terms of validity and reliability.

Predicting or substituting a value for the missing item response is called single imputation. On the other hand, the multiple imputation procedure replaces each missing item response with a set of plausible values.<sup>[1]</sup> Thus, uncertainty due to missing data can be addressed by the variation among a set of plausible values, and this cannot be achieved via single imputation. The multiple imputation procedure consists of three distinct steps: the imputation phase, the analysis phase, and the pooling phase. A proper method should be chosen in the imputation phase so that the specific context of the missing data problem can remain under consideration.

Different imputation methods have been developed for addressing the reasons that underlie the missing

data. However, not every method is suitable for every missing data problem, so it is important to know the type of study being conducted before deciding on an appropriate method. The response function (RF) method was chosen for the imputation phase of the multiple imputation process<sup>[2]</sup> because of the nature of the data in this study and this approach’s proven performance in recent simulation studies.<sup>[2-4]</sup>

Nowadays, with the help of computer technology, imputation techniques of varying complexity have become available and are being employed in many disciplines, including the health sciences. For example, rheumatology researchers have used imputation techniques to deal with the missing data problem in their studies<sup>[5,6]</sup> while others have investigated the missing data techniques themselves.<sup>[7-10]</sup> However, in medical studies, little has been done concerning the evaluation of the effect of missing data imputation methods on Rasch model person estimates.<sup>[11,12]</sup> In addition, no study that we know of has evaluated how the imputation of missing item responses in the Health Assessment Questionnaire Disability Index (HAQ-DI) data affects a patient’s disability measurements; therefore, the aim of this study was to accomplish this and determine how this imputation influences the bias and precision of the disability measurements.

---

## PATIENTS AND METHODS

---

### Patients and setting

Our data was collected from 389 outpatients with rheumatoid arthritis (RA) (n=174) and osteoarthritis (OA) (n=215) who had participated in previous studies performed at the Department of Physical Medicine and Rehabilitation, Ankara University, Faculty of Medicine between 2002 and 2011.<sup>[13-15]</sup> Out of the 389 patients, 385 (171 with RA and 214 with OA) had completed all of the items on the HAQ-DI. There were 309 females (80%) and 76 males (20%) in the study and the mean age and disease duration were 55±12 years (range; 18 to 84) and 8.5±8.3 years (range; 0 to 60), respectively. We used the responses of the 385 patients who had no missing item scores, and this led to a data set that had a complete set of answers for each respondent. After imposing missingness, multiple imputation and all analyses were performed using the scores of the eight sections of the HAQ-DI. All patients gave their informed consent to take part in this study, which was carried out in compliance with the Helsinki Declaration.

### Selected scale

The HAQ-DI is the most widely utilized self-report questionnaire to assess the functional status of patients with a variety of rheumatic diseases. After it was introduced in the 1980s for RA,<sup>[16]</sup> it is then applied to other diseases such as OA, juvenile RA, systemic lupus erythematosus (SLE), scleroderma, ankylosing spondylitis (AS), fibromyalgia, and psoriatic arthritis.<sup>[17]</sup> The HAQ-DI assessment instrument includes the eight domains of dressing and grooming, arising, eating, walking, hygiene, reach, grip, and common daily activities. For these eight domains, there are 20 questions with four possible responses (without any difficulty: 0, with some difficulty: 1, with much difficulty: 2, unable to do: 3). The highest score reported by the patient across any component question of the eight domains is recorded as the score of that domain unless aids or devices are required. In that case, the score is automatically raised to 2 when it is rated as 0 or 1. The HAQ-DI score is then calculated as the average of the eight domains (items) with scores ranging between 0 and 3, with a higher score representing more disability. The Turkish adaptation was used in the study.<sup>[18]</sup>

### Missing data simulation

Data sets with the missing item responses were created to evaluate the performance of the RF imputation technique with regard to patient disability level estimates. Item responses were deleted from the full data set ( $n=385$ ) with respect to the MCAR mechanism, and the missing data was generated through simple random selection from among all respondents with three missingness proportions (0.10, 0.30, and 0.50).

### Multiple imputation and Rasch model estimates

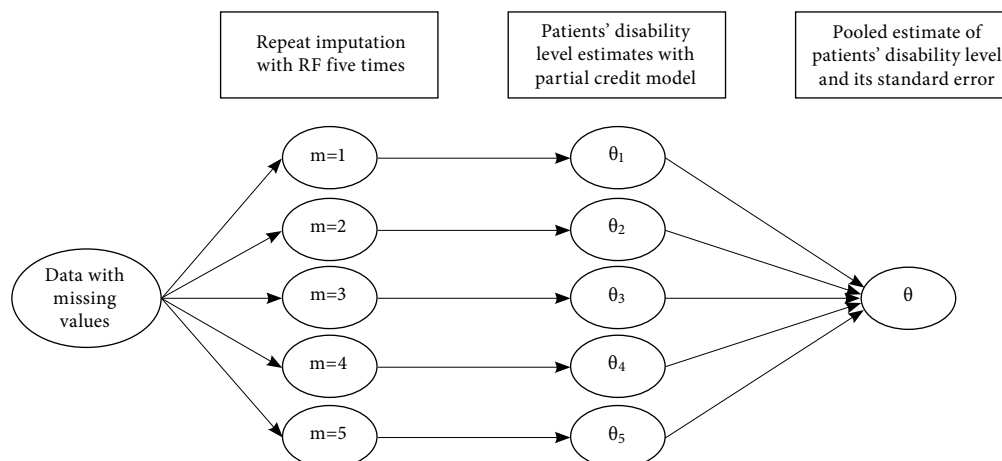
Multiple imputation was carried out separately for each of the three newly created data sets. In the imputation phase, the missing responses were imputed five times with different plausible values using the RF method. Then in the analysis phase, a Rasch model was used, and the patient disability levels were estimated for each of the completed data sets. In the pooling phase, the patient disability estimates and standard errors were combined into a single set of results for each of the three data sets (Figure 1).

### Imputation method

The RF imputation method was first proposed by Sijtsma and Van der Ark<sup>[2]</sup> for data related to test or scale. In the Rasch model, for a patient with a latent trait level, the probability of having a score  $x$  on item  $j$  is called the item response function, shown as,  $P(X_j=x|\theta)$ . The RF imputation uses the estimated item response function to impute item scores, and it has been proven to be an efficient imputation method for unidimensional scales in simulation studies.<sup>[2-4]</sup>

### The classical test theory and item response theory

The item response theory (IRT) is a modern test theory used for the design, analysis, and scoring of scales that are utilized to measure latent traits. It is generally considered to be superior to the classical test theory (CTT) due to its more cogent theoretical justifiability. In the IRT, the true score is defined by the latent trait level of interest ( $\theta$ ) rather than the ordinal raw score used in the CTT. The Rasch model, a one-parameter IRT model, helps to measure the



**Figure 1.** Graphic representation of the multiple imputation process for each of the newly created three data sets with missing values.  $\theta$ : a vector of the 385 patient disability estimates and standard errors.

latent trait levels of patients using the categorical response data collected to assess them.<sup>[19]</sup> Therefore, it has a specific property that provides a criterion for objective and successful measurement. Because the polytomous nature of the responses and the distance between thresholds across items were not similar, Master's partial credit model (PCM),<sup>[20]</sup> one of the Rasch models, was used to analyze the HAQ-DI data in this study.

To evaluate the similarity between the disability estimates from the original data and those from the multiple imputed data, a scatter plot and an ICC were used. We also evaluated the similarity between the disability estimates from the original data and those from the data with missing values. We found that the bias increased as the similarity between the two disability estimates was impaired. The same ICC calculations were also performed for the standard errors of disability estimates, and they were used for calculating between the items both before and after imputation.

The missing data simulation, multiple imputation with the RF method, pooling of the disability estimates and their standard errors, and scatter plots were performed using functions written in the R software package version 2.13.0 (The R Foundation for Statistical Computing).<sup>[21]</sup> Readily available functions in the extended Rasch modeling package (eRm) of R were used for PCM fit and the patient disability estimates.<sup>[22]</sup> The Statistical Package for the Social Sciences (SPSS Inc., Chicago, Illinois, USA) for Windows version 15.0 was used to calculate the ICC and its 95% confidence interval (CI). The R codes used in this study will be provided by the authors upon request with no expiration date (the RF imputation used in this study is also available as SPSS syntax that is freely downloadable from [<http://spitswww.uvt.nl/~avdrark/research/research.htm>]).<sup>[23]</sup>

## RESULTS

For the incomplete data case, the ICC between the disability estimates from the original data and the data with missing values diminished slightly as the proportion of the missingness increased. However, even when the missing data proportion was 0.50, the ICC was still close to one, and its 95% CI was still narrow (Table 1).

When the multiple imputed incomplete data sets using RF were calculated, the resulting ICCs were slightly lower than those found in cases with no imputation. In terms of standard errors, we found higher ICCs between the standard errors of disability estimates from the original data and the multiple imputed data sets, whereas the ICCs between the standard errors of disability estimates from the original data and the incomplete data sets were close to zero. The increased missingness proportion is associated with the diminished ICC (Table 1).

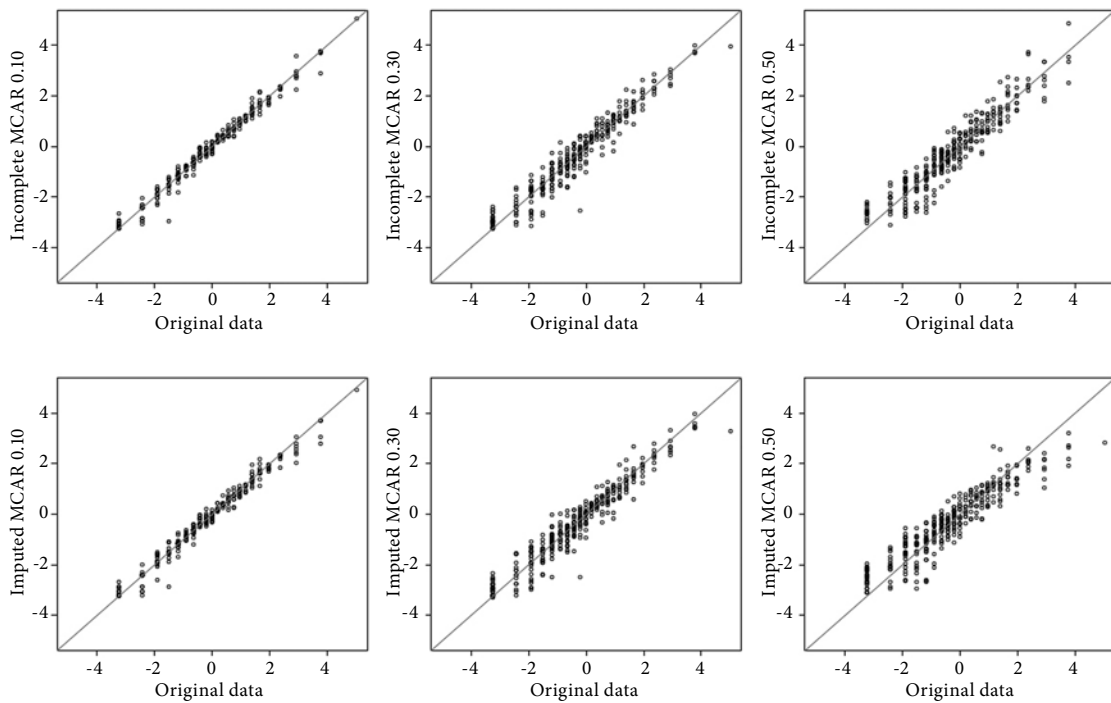
The scatter plots for disability estimates indicated that as the proportion of the missing values increased, the bias of disability estimates obtained from the incomplete data also increased. After the multiple imputed incomplete data set was examined, this bias was slightly larger than that of the incomplete data. Up to a proportion of 0.30 missingness, this bias was negligible, and even for 0.50 missingness, which only occurs in extreme cases, the disability estimates were close to those of the original data (Figure 2).

The standard errors of disability estimates were scattered over a wide range for the incomplete data when compared with the original data. The high proportion of missingness (0.50) caused these estimates to be slightly worse than the missingness proportions for 0.10 and 0.30. After the evaluation of the multiple

**Table 1.** Intra-class correlation coefficient between the original full data with missing values and the multiple imputed data for disability estimates along with the standard errors of disability estimates

	ICC and 95% CI for disability estimates	ICC and 95% CI for standard errors of disability estimates
Incomplete data MCAR 0.10	0.996 (0.995; 0.997)	0.016 (~0; 0.205)
Incomplete data MCAR 0.30	0.983 (0.979; 0.986)	~0 (~0; 0.139)
Incomplete data MCAR 0.50	0.970 (0.963; 0.976)	~0 (~0; ~0)
Multiple imputed data MCAR 0.10	0.995 (0.994; 0.996)	0.982 (0.978; 0.986)
Multiple imputed data MCAR 0.30	0.981 (0.977; 0.985)	0.947 (0.935; 0.957)
Multiple imputed data MCAR 0.50	0.958 (0.948; 0.966)	0.883 (0.855; 0.905)

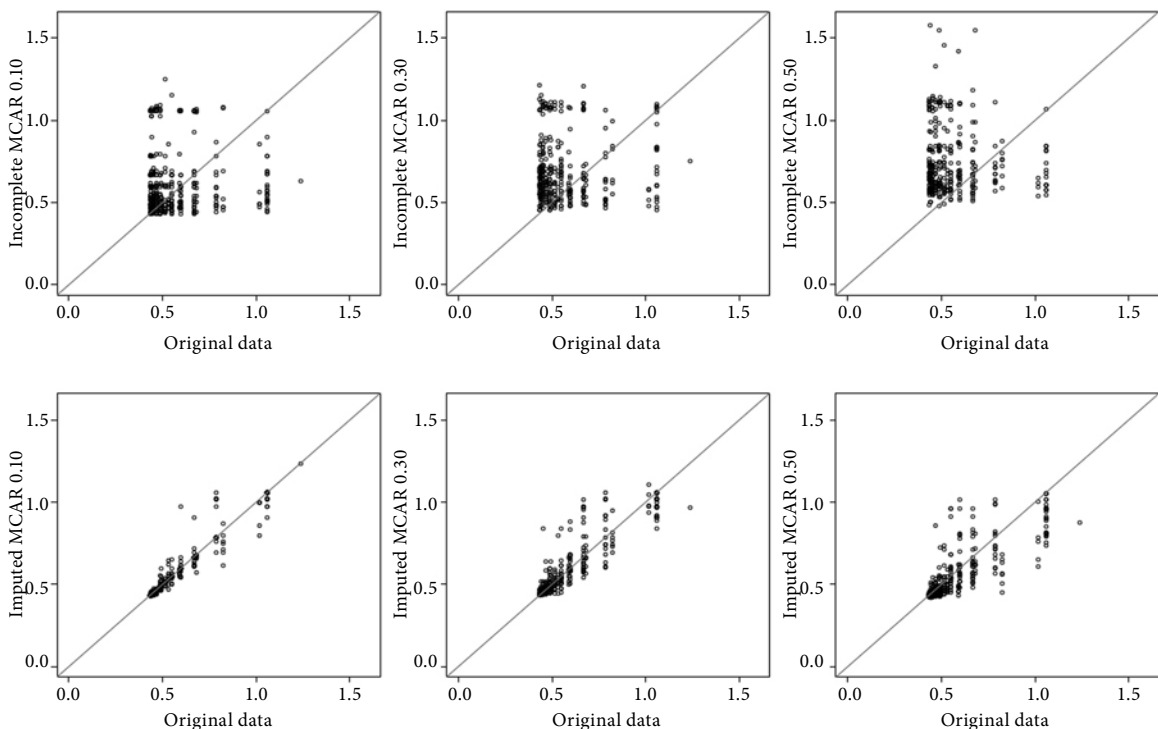
MCAR: Missing completely at random; ICC: Intra-class correlation coefficient; CI: Confidence interval; ~ Close to.



**Figure 2.** Scatter plots of disability estimates obtained from the original data, incomplete data, and the multiple imputed data. MCAR: Missing completely at random.

imputed incomplete data, the standard errors got closer to those of the original data, even with regard to the high proportion of missingness (Figure 3).

As for the ICC between items, the ICCs for multiple imputed data sets were closer to those of the original data than for those of the incomplete data sets. Also,



**Figure 3.** Scatter plots of standard errors of disability estimates obtained from the original data, incomplete data, and the multiple imputed data. MCAR: Missing completely at random.

**Table 2.** Intra-class correlation coefficient between the items and their related 95% CIs

	ICC	95% CI
The original full data (n=385)	0.892	0.875;0.908
Incomplete data MCAR 0.10 (n=162)	0.873	0.842;0.901
Incomplete data MCAR 0.30 (n=18)	0.915	0.841;0.963
Incomplete data MCAR 0.50 (n=2)	0.940	0.662;1.000
Multiple imputed data MCAR 0.10 (n=385)	0.890	0.873;0.906
Multiple imputed data MCAR 0.30 (n=385)	0.899	0.871;0.905
Multiple imputed data MCAR 0.50 (n=385)	0.875	0.855;0.893

MCAR: Missing completely at random; ICC: Intra-class correlation coefficient; CI: Confidence interval.

the 95% CIs of the ICCs for multiple imputed data sets were narrower than those for the data sets with missing values (Table 2).

## DISCUSSION

Missing data often occurs in self-reported outcomes, and this may affect the content validity, reliability, and power of the study and also have a negative effect on the standard error of disability estimate. Since the main goal of research and clinical assessment is to obtain precise, accurate, reliable, and valid results for the population of interest, the missing data presents a challenge for researchers that should not be ignored. Complete data is needed for item analysis within the CTT. This is in contrast to the IRT, which can handle missingness because the estimation of a respondent's latent trait level is based on the observed item responses. Therefore, imputation of missing values is feasible for both classical and modern test theories.

It is known that for data sets in accordance with the Rasch model, missing item responses do not cause bias on latent trait estimates, but they do lower the precision of the estimates and lessen the sensitivity of the fit statistics of the model. Thus, a multiple imputation strategy can be used to address the uncertainty caused by the missing data, which in turn improves the precision.

According to the results of this study, it can be concluded that both the disability estimates and their standard errors were affected by increasing the missing data proportion. However, the ICC values for the disability estimates stayed close to one. This showed that the bias in the disability estimates was negligible for cases with and without imputation. After the multiple imputation of the missing item responses by the RF method, the standard errors of estimates were found to be close to those of the original data.

De Ayala<sup>[11]</sup> reported that increasing the amount of missingness caused bias and higher standard errors for latent trait estimates. However, the missing data mechanism in his study was missing not at random (MNAR) in which the missingness is related to the missing item response itself. In contrast to De Ayala's<sup>[11]</sup> study, we found that increased missingness proportions caused bias in latent trait estimates. Our missing data was MCAR, so the difference in the results with respect to bias might arise from the different types of missing data mechanisms, number of items, and the different Rasch models used in the two studies.

In a recent study by Furlow et al.<sup>[12]</sup> the effect of missing data and differential item functioning on latent trait estimates from two polytomous Rasch models and different imputation methods (complete-case analysis, mean substitution, hot-decking, and multiple imputation based on multivariate normal) were compared using the MCAR mechanism, and they found that the presence of data associated with missingness increases the standard error of latent trait estimates but does not impact the bias in the theta estimates in the MCAR scenario. In our study, we used multiple imputation based on RF using the MCAR data, and our findings concur with the study by Furlow et al.<sup>[12]</sup>

In a study by Olsen et al.,<sup>[10]</sup> the Outcome Measures in Rheumatology - Osteoarthritis Research Society International (OMERACT-OARSI) set of responder criteria, the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) pain and physical function components from the WOMAC questionnaire, and a patient global assessment were used as outcome measures. The performance of five missing data handling methods (ignoring of the missing data, the last observation and the baseline observation being carried forward, and multiple imputation with two different strategies) were compared, and it was concluded that the multiple imputation of missing data

may be used to decrease bias and mean square error and increase power in OA trials.

Mirmohammadkhan et al.<sup>[9]</sup> compared multiple imputation with complete case analysis and found that it should be used to estimate the prevalence of rheumatic disorders like knee OA.

Our results are limited to the data collected via the HAQ-DI, an eight-item, four-point scale, and they are not generalizable to longer instruments or different types of Likert scales. The findings of this study are also limited to conditions which have missingness proportions of at most 0.50 under MCAR. Further studies could include simulations for the evaluation of the RF method along with other imputation methods with the HAQ-DI data using different missingness mechanisms.

In conclusion, imputed data sets and incomplete data sets showed almost the same bias in the disability estimates, whereas the precision of the disability estimates was better for the imputed data sets than for those of the incomplete data sets for each of the missing data proportions. When researchers encounter missing values in data collected with the HAQ-DI, multiple imputation with the RF method could be a convenient approach to overcome this dilemma.

#### Declaration of conflicting interests

The authors declared no conflicts of interest with respect to the authorship and/or publication of this article.

#### Funding

The authors received no financial support for the research and/or authorship of this article.

---

### REFERENCES

- Rubin DB. Multiple imputation for nonresponse in surveys. 2nd ed. New York: John Wiley & Sons; 1987.
- Sijtsma K, Van der Ark LA. Investigation and treatment of missing item scores in test and questionnaire data. *Multivar Behav Res* 2003;38:505-28.
- Van Ginkel JR, Van der Ark LA, Sijtsma K. Multiple imputation of item scores in test and questionnaire data, and influence on psychometric results. *Multivar Behav Res* 2007;42:387-414.
- Doğanay Erdoğan B. Assessing the performance of multiple imputation techniques for Rasch models with a simulation study. Ankara University, Graduate School of Health Sciences. PhD thesis; 2012.
- Wolfe F, Michaud K. The loss of health status in rheumatoid arthritis and the effect of biologic therapy: a longitudinal observational study. *Arthritis Res Ther* 2010;12:R35.
- Kavanaugh A, Fleischmann RM, Emery P, Kupper H, Redden L, Guerette B, et al. Clinical, functional and radiographic consequences of achieving stable low disease activity and remission with adalimumab plus methotrexate or methotrexate alone in early rheumatoid arthritis: 26-week results from the randomised, controlled OPTIMA study. *Ann Rheum Dis* 2012. [Epub ahead of print]
- Jenkinson C, Heffernan C, Doll H, Fitzpatrick R. The Parkinson's Disease Questionnaire (PDQ-39): evidence for a method of imputing missing data. *Age Ageing* 2006;35:497-502.
- Wong WK, Boscardin WJ, Postlethwaite AE, Furst DE. Handling missing data issues in clinical trials for rheumatic diseases. *Contemp Clin Trials* 2011;32:1-9.
- Mirmohammadkhan M, Foroushani AR, Davatchi F, Mohammad K, Jamshidi A, Banihashemi AT, et al. Multiple Imputation to Deal with Missing Clinical Data in Rheumatologic Surveys: an Application in the WHO-ILAR COPCORD Study in Iran. *Iran J Public Health* 2012;41:87-95.
- Olsen IC, Kvien TK, Uhlig T. Consequences of handling missing data for treatment response in osteoarthritis: a simulation study. *Osteoarthritis Cartilage* 2012;20:822-8. doi: 10.1016/j.joca.2012.03.005.
- De Ayala RJ. The effect of missing data on estimating a respondent's location using ratings data. *J Appl Meas* 2003;4:1-9.
- Furlow CF, Fouladi RT, Gagne P, Whittaker TA. A Monte Carlo study of the impact of missing data and differential item functioning on theta estimates from two polytomous Rasch family models. *J Appl Meas* 2007;8:388-403.
- Öztuna D. An application of computerized adaptive testing in the evaluation of disability in musculoskeletal disorders. Ankara University, Graduate School of Health Sciences. PhD thesis; 2008.
- Elhan AH. Studies of Rasch analysis and its application on the physical medicine and rehabilitation data. Ankara University, Graduate School of Health Sciences. PhD thesis; 2002.
- Kaskatı OT. Development of computer adaptive testing method using Rasch models for assessment of disability in rheumatoid arthritis patients. Ankara University, Graduate School of Health Sciences. PhD thesis; 2011.
- Fries JF, Spitz P, Kraines RG, Holman HR. Measurement of patient outcome in arthritis. *Arthritis Rheum* 1980;23:137-45.
- Bruce B, Fries JF. The Stanford Health Assessment Questionnaire: a review of its history, issues, progress, and documentation. *J Rheumatol* 2003;30:167-78.
- Küçükdeveci AA, Sahin H, Ataman S, Griffiths B, Tennant A. Issues in cross-cultural validity: example from the adaptation, reliability, and validity testing of a Turkish version of the Stanford Health Assessment Questionnaire. *Arthritis Rheum* 2004;51:14-9.
- Pallant JF, Tennant A. An introduction to the Rasch measurement model: an example using the Hospital Anxiety and Depression Scale (HADS). *Br J Clin Psychol* 2007;46:1-18.

20. Masters GN. A rasch model for partial credit scoring. *Psychometrika* 1982;47:149-74.
21. R Development Core Team, R: A language and environment for statistical computing, a foundation for statistical computing, Vienna, Austria: ISBN 3-900051-07-0. Available from: <http://www.R-project.org>. [Date of access 2011]
22. Mair P, Hatzinger R, Maier MJ. eRm: Extended rasch modeling. R package version 0.15-0. Available from: <http://CRAN.R-project.org/package=eRm> [Date of access 2012]
23. Van Ginkel JR, Van der Ark LA. SPSS syntax for missing value imputation in test and questionnaire data. *Appl Psych Meas* 2005;29:152-3.