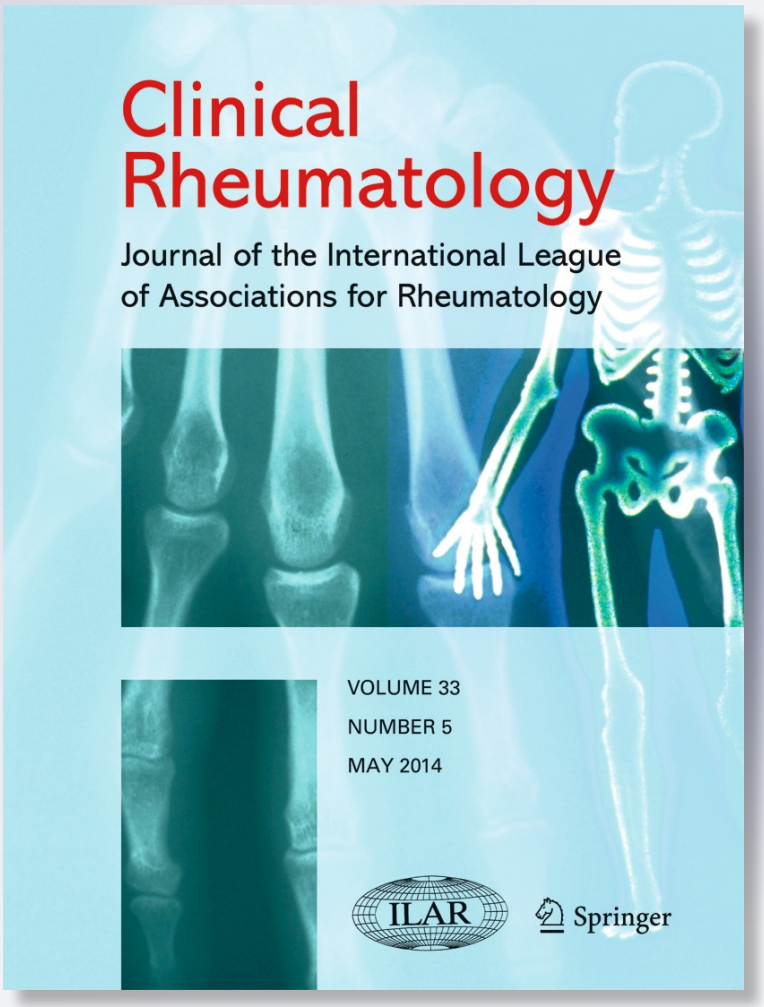
Assessment of the dimensions, construct validity, and utility for rheumatoid arthritis screening of the COPCORD instrument

José Moreno-Montoya, José Alvarez-Nemegyei, Belem Trejo-Valdivia & Ingris Peláez-Ballestas

Clinical Rheumatology
Journal of the International League of Associations for Rheumatology

ISSN 0770-3198
Volume 33
Number 5

Clin Rheumatol (2014) 33:631-636
DOI 10.1007/s10067-013-2367-3



Springer

Clin Rheumatol (2014) 33:631–636
DOI 10.1007/s10067-013-2367-3

ORIGINAL ARTICLE

Assessment of the dimensions, construct validity, and utility for rheumatoid arthritis screening of the COPCORD instrument

José Moreno-Montoya · José Alvarez-Nemegyei · Belem Trejo-Valdivia · Ingris Peláez-Ballestas

Received: 17 January 2013 / Revised: 29 July 2013 / Accepted: 8 August 2013 / Published online: 24 August 2013

Abstract This study aims to evaluate the structural validity of the Community-Oriented Program for the Control of Rheumatic Diseases (COPCORD) core instrument as a screening tool for rheumatoid arthritis (RA) by means of assessing the existence of domains in the questionnaire. The Mexican version of the COPCORD instrument was applied to individuals over18 years of age in five regions of the country through a probabilistic/convenience household survey. Clinical confirmation of RA diagnosis was used. The variables analyzed included self-reported comorbidities and manifestations of the disease, as well as sociodemographic characteristics. The statistical approach was based on polychoric exploratory factor analysis and confirmatory factor analysis by means of probit structural equation models. A total of 19,213 subjects were included in the analysis. The average age for the total sample was 42.89 years old; 40.64 % of the subjects were older than 45 years of age and 20.42 % older than 55. More than 80 % of the variation was related to three underlying factors: recent pain, historical pain, and disability. The findings verified the usefulness of the COPCORD instrument as a screening tool for RA. The results also allowed to characterize how the variation in terms of manifestations of the disease could be accounted for diagnosing the disease in the Mexican context and examined the capabilities of the instrument to measure correctly the main characteristics of patients suffering from RA.

Keywords Community · COPCORD · Factor analysis · Rheumatoid arthritis · Screening tool

Introduction

Rheumatic diseases are the main cause of disability worldwide, producing large personal, family, and economic burdens [1]. This situation has motivated efforts at the international level, including the declaration of the World Health Organization's (WHO) "Bone Decade" [2]. In the late 1980s, the International League Against Rheumatic Diseases (ILAR), together with the WHO, launched the Community-Oriented Program for the Control of Rheumatic Diseases (COPCORD), which aimed to control and reduce the burden of the disease. This initiative, designed for societies with limited access to healthcare systems, includes screening for rheumatic diseases based on the identification of musculoskeletal (MSK) complaints, in particular pain/swelling/stiffness and restricted range of motion in the joints and/or MSK soft tissue [3]. Nowadays, several countries have developed different epidemiological approaches to the local validation of the COPCORD instrument, as well as the required cultural adaptation [4, 5]. In Mexico, a five-state study was developed where around 20,000 adults were examined by means of the

GEEMA (Grupo de Estudio Epidemiológico de Enfermedades Musculo Articulares): R. Burgos-Vargas, Hospital General de México, Faculty of Medicine, Universidad Nacional Autónoma de México, México City; MH Cardiel, Hospital Secretaría de Salud, Hospital General "Dr. Miguel Silva."; L.H. Sanín, Universidad Autónoma de Chihuahua, Instituto Nacional de Salud Pública; M. Garza-Elizondo, Hospital Universitario José Eleuterio González; J. Rodriguez-Amado, Hospital Universitario José Eleuterio González; M.V. Goycochea-Robles, Unidad de Investigación, Colegio Mexicano de Reumatología; M. Madariaga, Hospital General de Culiacán, Secretaría de Salud; J. Zamudio, Hospital General de Culiacán, Secretaría de Salud; Natalia Santana Portillo, Instituto Mexicano del Seguro Social, Chihuahua.

J. Moreno-Montoya · B. Trejo-Valdivia
Instituto Nacional de Salud Pública, Cuernavaca, Mexico

J. Alvarez-Nemegyei
Universidad Anahuac-Mayab, Mérida, Mexico

I. Peláez-Ballestas (✉)
Rheumatology Department, Hospital General de México, Dr. Balmis 148, Col. Doctores, Delegación Cuauhtémoc CP 06720 Mexico City, Mexico
e-mail: pelaezin@prodigy.net.mx

instrument and then diagnosed [6]. This study also allowed the identification of important features of the instrument's operative characteristics [3], providing information about its performance as a classification tool.

The objective of our research was to evaluate the structural validity of the COPCORD core instrument. This analysis, carried out through factorial analysis at exploratory and confirmatory levels, allowed us to assess the capabilities of the instrument to measure correctly the main characteristics of patients suffering from rheumatoid arthritis (RA) and, as a consequence, contribute to the characterization of its performance as a screening tool.

Methods

Study population

A detailed description of the COPCORD survey in Mexico has been reported elsewhere [6]. Briefly, a version of the COPCORD instrument validated for the Mexican population [5] was applied to 19,213 individuals over 18 years of age—11,602 (60.39 %) women and 7,611 (39.61 %) men—who participated in a study on the prevalence of rheumatic diseases in Mexico. The study was based on a cross-sectional community survey in 24 municipalities distributed throughout in five states of Mexico: northern Mexico (Chihuahua), and the Northeast (Nuevo León), Northwest (Sinaloa), the South (Yucatán), and the center of the country (Mexico City). The sample was drawn by using a multistage random probability scheme across four states and a convenience sample for Mexico City. The 2005 population census, stratified by region, was used as a sampling frame. Municipalities were chosen according to a proportional sample size design, and finally a door-to-door survey was carried out using the WHO/ILAR COPCORD stage 1 questionnaire [6]. All individuals reporting joint pain, swelling, or stiffness over the previous 7 days and previously in their lives were evaluated by board-certified rheumatologists, and RA diagnoses were based on standardized medical classification criteria according to the 1987 American College of Rheumatology [7]. The overall prevalence of RA was 1.6 % (95 % confidence interval (CI) 1.42–1.78), with 2.09 % (95 % CI 1.83–2.35) in women and 0.85 % (95 % CI 0.64–1.06) in men.

COPCORD questionnaire

The COPCORD core questionnaire includes self-reports on comorbidities, pain manifestations, use of addictive substances, and disability [3]. Specifically, the variables involved are age, sex, self-reports on smoking, drug use, and alcohol use (Y/N), and "yes or no" self-reports on diabetes mellitus, high blood pressure, heart disease, peripheral vascular disease, gastritis, anxiety, depression, obesity, and hyperlipidemia. To obtain information about the clinical manifestations of rheumatologic diseases, the instrument includes questions related to pain and trauma in the more distant past as well as during the previous 7 days. Disability was measured using the validated Mexican version of the Health Assessment Questionnaire (HAQ), which includes questions about whether the patients can carry out common activities such as getting down on their knees [8]. Those interviewed were also asked about previous medical treatments and diagnoses for any diseases other than RA. Some sociodemographic characteristics are also included, such as income (less than US$192 per month), education (years of formal education completed), type of job, whether or not they handle heavy loads, age at the time the survey was completed, and gender. Patients reporting pain because of recent traumas were excluded.

Data analysis

Two statistical approaches were used in this study: an exploratory factor analysis (EFA) and a confirmatory factor analysis (CFA). The major goal of both was to identify and model the relationships among the observed variables, using a smaller number of unobserved or latent factors [9]. For the EFA, nothing was hypothesized about the structure or number of factors in the COPCORD instrument or even how each observed variable would relate to any other. The correlations were calculated by means of polychoric coefficients [10]. For the CFA, the factors used were those identified in the previous step. The confirmatory process was developed using a robust weighted least squares (WLS) estimation method to reduce the effect of assuming a continuous normal latent process underlying the non-normal/categorical variables in the instrument [11]. Model diagnostics were developed using the Satorra–Bentler scaled χ^2 as well as robust standard errors [11, 12].

After carrying out the CFA, a structural equation model was built, using the result of the screening for RA as the outcome variable. Details about the diagnosis considerations and procedures have been described previously [3]. The model had two goals: first, to provide information about the domain structure of the instrument regarding the RA diagnosis and, second, to construct a pathway model to facilitate the understanding on the relationships among the variables linked to instrument domains. A multiple-stage WLS probit regression was used [13]. Because of the size of the calculations involved, the model was made using the values obtained from the CFA as initial values in the path [14].

The steps for building the model were as follows: first, the construction of a full model, involving all the variables and, second, an assessment of the modification indices provided by the Linear Structural Relations Package (LISREL) giving information about which variables were most important in

Table 1 Total variance explained and eigenvalues

Factor	Eigenvalues	Variance explained (%)
1	5.01	54.97
2	3.79	16.56
3	1.54	8.53
4	0.57	9.90
5	0.46	7.99

The figures indicate that there are only three significant factors (they associated to eigenvalues bigger than 1), and that they are able to explain above 80 % of the observed variation (once adding the explained variance from each one)

the model, as well as the correlations between variables, taking into account underlying latent characteristics, for instance, disability.

The correlated errors in the resulting model were used to search for further relationships among the variables, especially those measuring the same aspects of any dimension (for instance, historical pain and pain during the previous 7 days).

The analyses were developed using the LISREL package [15] for EFA and polychoric matrix calculations. The Stata 12 package was used for the confirmatory analysis of principal components [16]. The structural model was created using the Structural Equation Modelling module of R software [17] and the Conditional Mixed Process Estimator macro for Stata 12 by Roodman [18].

Table 2 Total variance explained and eigenvalues

Item/question	Loading[a]	Factor	Factor issue
Pain during the last 7 days	0.7862	1	Related with recent pain
Intensity of the pain during the last 7 days	0.8046	1	
Previous medical treatments	0.7333	1	
Historical pain	0.7596	2	Related with historical pain
Intensity of historical pain	0.7488	2	
Difficulty for dressing	0.6907	3	Related with disability
Difficulty for lying down	0.7523	3	
Difficulty for getting up	0.5828	3	
Difficulty for walking	0.6551	3	
Difficulty for washing	0.6845	3	
Difficulty for bowing	0.7321	3	
Difficulty for opening with keys	0.6196	3	
Difficulty for crouching	0.7233	3	
Difficulty for kneeling	0.7654	3	

The loading values in this case indicate the mean input for every item inside each factor. This way, each variable underlying or unobserved (latent) could be mathematically obtained by adding these loadings once each single variable is measured

[a] These values are obtained by means of exploratory analysis and a confirmatory one is needed. The confirmatory analysis was also developed by means of structural equation models and is shown in Table 3

Table 3 CFA linear models for each factor

Factor/variable	Coef.	Std. Error	Z value	$p>Z$	95 % CI	
Factor 1						
Intensity of the pain during the last 7 days	0.899	0.001	19.44	0	0.898	0.900
Previous medical treatments	0.774	0.004	214.8	0	0.768	0.782
Constant	0.051	0.002	29.03	0	0.048	0.055
Factor 2						
Intensity of historical pain	0.847	0	24.71	0	0.846	0.848
Constant	0.028	0.001	27.14	0	0.026	0.030
Factor 3						
Difficulty for dressing	0.715	0.006	18.12	0	0.704	0.726
Difficulty for lying down	0.734	0.005	14.31	0	0.725	0.743
Difficulty for getting up	0.585	0.008	73.73	0	0.570	0.601
Difficulty for walking	0.678	0.005	18.24	0	0.668	0.687
Difficulty for washing	0.67	0.007	10.56	0	0.657	0.683
Difficulty for bowing	0.775	0.004	19.09	0	0.767	0.783
Difficulty for opening with keys	0.623	0.007	93.09	0	0.610	0.636
Difficulty for crouching	0.751	0.005	15.25	0	0.741	0.760
Difficulty for kneeling	1.414	0.003	32.26	0	1.408	1.421
Constant	0.017	0.001	15.83	0	0.015	0.019

The coefficients and every related p value and confidence interval show the "corrected" estimations of the loadings of each item inside its correspondent factor. This correction is obtained by means of the structural equation model (CONFA) and incorporates a constant term designed to improve the adjusting of the factors

Results

A total of 19,213 subjects were included in the analysis. The average age for the total sample was 42.89 years old (SD

Table 4 Structural equations model

Latent variables	Coef.	Std. Error	Z value	$p>Z$	95 % CI	
Factor 1	0.037	0.009	4.01	0	0.019	0.055
Factor 2	0.068	0.009	7.32	0	0.050	0.086
Factor 3	0.082	0.024	3.39	0.001	0.035	0.130
Constant	−2.898	0.102	−28.53	0	−3.097	−2.699

The coefficients and its associated p values and confidence intervals show the participation of each factor (composed for the variables/items shown in Table 3) in the likelihood of being classified as with a positive diagnosis of RA. This way, this model represents a complex system of linear regressions where each factor is at the same time a regression model. All variables, observed (items) and latent (factors), are indeed related with the RA outcome, but its participation is mediated and controlled for the kind of information every variable is giving. In the case of each item, providing specific information about the correspondent domain (for instance, different traits of disability in factor 3), and in the case of each factor, providing information of a specific construct underlying the manifestations of RA

17.34; min, 17; max, 99); 40.64 % of the subjects were older than 45 years of age and 20.42 % older than 55.

The principal axis factoring resulted in a three-factor solution (factors showing eigenvalues bigger than 1) (see Table 1). Cumulatively, the three factors account for 80.06 % of the variation. Loadings and variables for each factor are shown in Table 2. The higher item correlation was 0.53, and the model showed significant covariation structure [$p > \chi^2 = 0.000$]. The CFA model showed similar values to the initial loadings identified in the EFA model. All the variables included showed significant coefficients within each factor. However, two variables, "pain during the previous 7 days" and "historical pain," were excluded from the model because of colinearity. The linear models related to each factor are shown in Table 3.

This procedure achieved at least one significant coefficient within each factor in the structural equation model. The results were stable whether or not the variables of age, region, sex, and income were included. The variables related to use of addictive substances seem to be unrelated to the diagnosis of RA. The coefficients and standard errors and confidence intervals are shown in Table 4. A graphical representation of the model is shown in Fig. 1, where the latent variables or factors are indicated by circles and the observed variables, by squares. The observed exogenous variables are labeled by name. The observed endogenous variable is labeled as RA. The paths are indicated by the value of the coefficient in the model. The errors are also exposed. For the latent endogenous variables, the errors are labeled psi (ξ). For the observed exogenous variables, these errors are labeled delta (δ). Residual covariance among latent variables is labeled phi (φ).

Discussion

This study explored the structural validity and construct properties of the COPCORD instrument. The factorial scheme showed that three factors could be extracted from the questionnaire. Estimations for specific sex–age–region grouping were stable and consistent. The total variance explained by the model was greater than 80 %; thus, the instrument is useful in

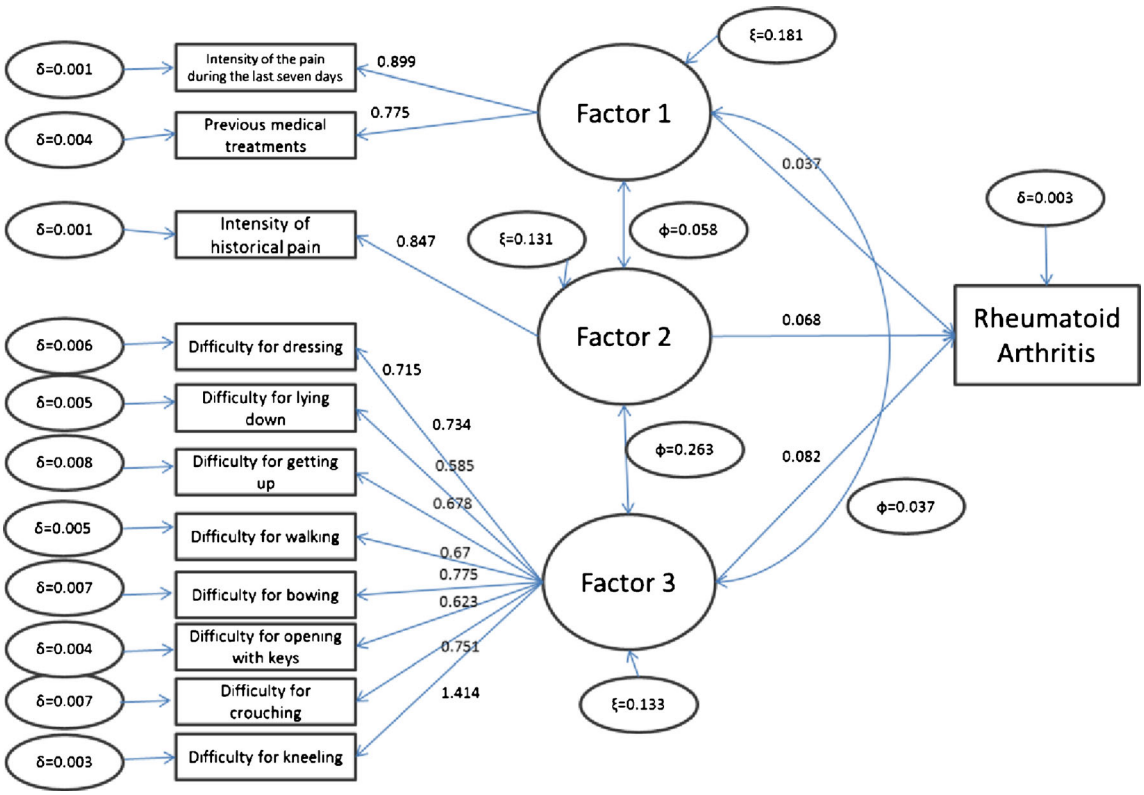


Fig. 1 Graphic representation of the system of structural equations. The complex structure of relationships underlying the diagnosis of RA in terms of the items composing the COPCORD instrument is shown. Every item, represented by a *rectangle*, has a measurement error δ and when measured, represents a part of the complete manifestation of a domain (factor). This is the reason why the *arrows* come from the latent variables to the observed ones. Instead, the latent variables (unobservable) are affecting directly the outcome (RA) but show also complex patterns of variation and covariation (characterized by the *bidirectional arrows* among factors and their unobservable variations ξ). Finally, the *arrows* coming from each latent variable and directed to the outcome variable try to represent how, once corrected by the complex pattern of covariation, each domain participates in the presence or not of the disease and its measurable manifestation

explaining the variance in the diagnosis of RA, and its structure is stable across genders and age groups.

Our findings point to the existence of three domains of information in the questionnaire or, mathematically, three latent variables. The first factor seems to be related to recent manifestations of pain in rheumatic patients, verifying the results reported elsewhere about the quality of life of these people [19]. The second factor consists of the variables measuring intensity and the presence of historical pain [20]. The third factor is related to disability. We used each HAQ item separately instead of the aggregated result of the HAQ scale, so as to take advantage of all the information provided by the disability questions and to avoid redundancy in the analysis linked to the result of the scale.

The system of equations developed excluded some variables because of colinearity. This means that some of the questions included in the COPCORD instrument could be redundant, for instance, about the presence and intensity of musculoskeletal pain during the previous 7 days. Similar results were obtained for the second factor, where one of the two highly correlated variables of historical pain and its intensity had to be removed from the model.

Our findings verify the existence of a clear pattern of covariation of the COPCORD instrument and support its use as a screening tool. Our results allow a better understanding about the symptoms and manifestations characterizing and grouping cases of RA. However, our model is unable to draw any conclusions about the risk or probability, or even relative risk, of developing RA. This exercise simply helps to clarify that the questionnaire is able to identify correctly sources of variation involved in the screening process and discriminate between patients regarding their characteristics and clinical manifestations.

The path model does not establish any causal relationships, although this was the first objective in the epidemiological literature [21]. Our cross-sectional design also fails to allow us to establish any causal relationship. Our findings must therefore be understood as validating which symptoms related to RA could happen before or after diagnosis and how they are related. We examined the effect of regional variation on the stability of the model and did not find significant changes in the coefficients. However, place-specific research is necessary to identify whether any particular covariation patterns exist.

Our results confirm the structural validity of the COPCORD instrument in supporting a specific diagnosis of RA and constitute a more formal approach than that published before [3]. In this work, we concentrated on the mathematical verification of the structure of the questionnaire and used that to identify the constructs underlying the variables in the instrument and their relationship with the clinical manifestations of the disease.

Acknowledgments This project was supported by the funds provided by the Salud 2007-C01-69439 and Salud 2007-C01 69765 programs of the National Council of Science and Technology of México (CONACYT) and the Fundación Mexicana para Enfermos Reumáticos (FUMERAC).

Disclosures None.

References

1. Toledano E, Candelas G, Rosales Z et al (2012) A meta-analysis of mortality in rheumatic diseases. Reumatol Clin 8(6):334–341. doi:10.1016/j.reuma.2012.05.006
2. Lidgren L (2012) Looking back at the start of the bone and joint decade what have we learnt? Best Pract Res Clin Rheumatol 26(2):169–171
3. Goycochea-Robles MV, Sanin LH, Moreno-Montoya J et al (2011) Validity of the COPCORD core questionnaire as a classification tool for rheumatic diseases. J Rheumatol 38(suppl 86):31–35
4. Haq SA, Davatchi F, Dahaghin S et al (2010) Development of a questionnaire for identification of the risk factors for osteoarthritis of the knees in developing countries. A pilot study in Iran and Bangladesh. An ILAR-COPCORD phase III study. Int J Rheum Dis 13(3):203–214
5. Bennett K, Cardiel MH, Ferraz MB et al (1997) Cross cultural adaptation and screening characteristics of the COPCORD Core Questionnaire in Brazil, Chile, and Mexico. J Rheumatol 24:160–168
6. Peláez-Ballestas I, Sanin LH, Moreno-Montoya J et al (2011) Epidemiology of the rheumatic diseases in Mexico. A study of 5 regions based on the COPCORD methodology. J Rheumatol Suppl 86:3–8
7. Alamanos Y, Voulgari PV, Drosos AA (2006) Incidence and prevalence of rheumatoid arthritis, based on the 1987 American College of Rheumatology criteria: a systematic review. Semin Arthritis Rheum 36:182–188
8. Álvarez-Hernández E, Peláez-Ballestas I, Vázquez-Mellado J et al (2008) Validation of the Health Assessment Questionnaire disability index in patients with gout. Arthritis Rheum 59(5):665–669
9. Flora BD, Labrish C, Chalmers P (2012) Old and new ideas for data screening and assumption testing for exploratory and confirmatory factor analysis. Front Psychol 3:55–57
10. Foreo CG, Maydeu-Olivares A, Gallardo-Pujol D (2009) Factor analysis with ordinal indicators: a Monte Carlo study comparing DWLS and ULS estimation. Struct Equ Model A Multidiscip J 16(4):625–641
11. David BF, Curran PJ (2004) An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. Psychol Methods 9(4):466–491
12. Satorra A, Bentler PM (1994) Corrections to test statistics and standard errors in covariance structure analysis. In: Von Eye A, Clogg C (eds) Latent variables analysis: applications for developmental research. SAGE, Thousand Oaks, pp 399–419
13. Yuan KH, Bentler PM (1998) Structural equation modeling with robust covariances. Sociol Methodol 28:363–396
14. Wirth RJ, Edwards MC (2007) Item factor analysis: current approaches and future directions. Psychol Methods 12:58–79
15. Scientific Software International, Inc (2012) LISREL: a general computer program for estimating a linear structural equation system involving multiple indicators of unmeasured variables (computer program). Scientific Software International, Inc., Chicago
16. StataCorp LP (2011) Stata statistical software (computer program). StataCorp, College Station
17. R Foundation for Statistical Computing (2012) R for Windows (computer program). R Foundation for Statistical Computing, Vienna

18. Roodman D (2011) Estimating fully observed recursive mixed-process models with cmp. Stata J 11(2):159–206
19. Uhm DC, Nam ES, Lee HY et al (2012) Health-related quality of life in Korean patients with rheumatoid arthritis: association with pain, disease activity, disability in activities of daily living and depression. J Korean Acad Nurs 42(3): 434–442
20. Lesuis N, Befrits R, Nyberg F et al (2012) Gender and the treatment of immune-mediated chronic inflammatory diseases: rheumatoid arthritis, inflammatory bowel disease and psoriasis: an observational study. BMC Med 10:82
21. Blood EA, Cheng DM (2011) The use of mixed models for the analysis of mediated data with time-dependent predictors. J Environ Public Health 2011:435078. doi:10.1155/2011/435078