Predictors of health-related quality of life in musculoskeletal disease patients: a longitudinal analysis

Alfredo Madrid-García (ID), **Leticia León-Mateos, Esperanza Pato, Juan A Jover, Benjamín Fernández-Gutiérrez** (ID), **Lydia Abasolo, Ernestina Menasalvas* and Luis Rodríguez-Rodríguez*** (ID)

Abstract

Introduction: Rheumatic and musculoskeletal diseases (RMDs) have a significant impact on patients' health-related quality of life (HRQoL) exacerbating disability, reducing independence and work capacity, among others. Predictors' identification affecting HRQoL could help to place efforts that minimize the deleterious impact of these conditions on patients' wellbeing. This study evaluates the influence of demographic and clinical predictors on the HRQoL of a cohort of RMD patients, measured using the Rosser classification index (RCI).

Methods: We included patients attending the Hospital Clínico San Carlos (HCSC) rheumatology outpatient clinic from 1 April 2007 to 30 November 2017. The primary outcome was the HRQoL assessed in each of the patient's visits using the RCI. Demographic and clinical variables extracted from a departmental electronic health record (EHR) were used as predictors: RMD diagnoses, treatments, comorbidities, and averaged HRQoL values from previous periods (for this last variable, values were imputed if no information was available). Association between predictors and HRQoL was analyzed using penalized generalized estimating equations (PGEEs). To account for imputation bias, the PGEE model was repeated excluding averaged HRQoL predictors, and common predictors were considered.

Discussion: A total of 18,187 outpatients with 95,960 visits were included. From 410 initial predictors, 19 were independently associated with patients' HRQoL in both PGEE models. Chronic kidney disease (CKD), an episode of prescription of third level analgesics, monoarthritis, and fibromyalgia diagnoses were associated with worse HRQoL. Conversely, the prescription in the previous visit of acid-lowering medication, colchicine, and third level analgesics was associated with better HRQoL.

Conclusion: We have identified several diagnoses, treatments, and comorbidities independently associated with HRQoL in a cohort of outpatients attending a rheumatology clinic.

Keywords: musculoskeletal disease, outpatient care, penalized regression methods, quality of life, risk factor

Correspondence to:
Luis Rodríguez-Rodríguez
Grupo de Patología Musculoesquelética, Hospital Clínico San Carlos, c\ Prof. Martin Lagos s/n, Madrid 28040, Spain
**lrrodriguez@salud
.madrid.org**

Alfredo Madrid-García
Grupo de Patología Musculoesquelética, Hospital Clínico San Carlos, IdISSC, Madrid, Spain

Escuela Técnica Superior de Ingenieros de Telecomunicación (ETSIT), Universidad Politécnica de Madrid, Madrid, Spain

**Leticia León-Mateos
Lydia Abasolo**
Grupo de Patología Musculoesquelética, Hospital Clínico San Carlos, IdISSC, Madrid, Spain

**Esperanza Pato
Juan A Jover
Benjamín Fernández-Gutiérrez**
Rheumatology Service, Hospital Clínico San Carlos, IdISSC, Madrid, Spain

Ernestina Menasalvas
Centro de Tecnología Biomédica, Universidad Politécnica de Madrid, Parque Científico y Tecnológico de la UPM, Madrid, Spain

*These authors shared senior authorship of this paper

Introduction

Health-related quality of life (HRQoL) assessment is a useful and well-extended way to measure, compare, and evaluate the overall (i.e. physical, social and mental) wellbeing of patients who suffer from chronic conditions, including those with rheumatic and musculoskeletal diseases (RMDs).[1–3] Hence, HRQoL is a measure of the patient's perception of his/her health status that comprises both subjective and objective components.[4]

According to the 2017 Global Burden of Disease report,[5] together with mental health disorders, RMDs were the leading cause of disability worldwide, affecting all age segments, and having the greatest negative impact on patients' HRQoL compared to other chronic disorders.[6,7] The high impact on disability caused by RMDs can be explained due to their high prevalence and incidence, affecting around 25% of the overall EU population.[8] Besides disability, other characteristics of RMDs such as chronicity, pain, fatigue, motion dysfunction, and multimorbidity[9] have a deleterious effect on the patients' HRQoL.

Different generic questionnaires for measuring the HRQoL have been employed in patients with chronic diseases,[10] as well as in RMD patients,[11,12] such as the Short Form 36,[13] EuroQol5D,[14] or the Rosser classification index (RCI).[15] Other RMD-specific questionnaires have also been proposed to enhance patients' characterization for specific disorders.[11,16]

Besides the increasing use of HRQoL in both clinical and observational studies as endpoints;[17] its relevance in the elaboration of medical guidelines and recommendations,[18] and in the design of medical interventions; attempts to identify predictors of HRQoL have also been made for different RMDs, such as Behçet's disease,[19] fibromyalgia,[20] gout,[21] juvenile idiopathic arthritis,[22] rheumatoid arthritis,[23] systemic lupus erythematosus,[24] and systemic sclerosis.[25]

However, most of the studies focus on finding HRQoL predictors in patients suffering a particular rheumatic disease, underestimating that the coexistence of other RMDs can have a synergistic effect on the subject's HRQoL.[3]

Therefore, we aimed to identify factors associated with the HRQoL of RMD patients attending an outpatient rheumatology clinic from a tertiary care center by considering the coexistence of multiple RMDs, the prescription of different medications and the presence of comorbidities, all of them potentially impairing the patient's HRQoL.

Patients and methods

Patients
The Hospital Clínico San Carlos Musculoskeletal Cohort (HCSC-MSKC)[26] is a routine clinical practice cohort that includes subjects seen at the rheumatology outpatient clinic of our center, whose clinical and management information is stored using a departmental electronic health record (EHR). The HCSC-MSKC contains information from more than 35,000 patients attending our clinic from 1 April 2007 until 30 November 2017.

In every patient's visit, information is collected both codified and as free text. The former includes RMD diagnoses, prescribed drugs, HRQoL (using the RCI), and the patient's follow-up plan. The latter contains the clinical notes, the comorbidity, and the medication prescribed by other physicians.

In the present study, we have selected patients and patients' visits from the HCSC-MSKC based on the following inclusion criteria:

- Patients' visits had to be spaced $\leqslant 365$ days apart from the previous and/or the following visit. Hence, the elapsed time between the first and the second visit and between the last and the previous visit had to be $\leqslant 365$ days. Isolated visits (i.e. visits separated by more than a year from the previous and the next visit) were excluded.
- Patients had to be seen at least twice in our outpatient clinic (and both visits had to fulfill the previous criteria).

Cognitive impairment or communication limitations were not exclusion criteria.

After applying these criteria, 18,187 patients and 95,960 visits were studied (Figure 1).

Based on the 365 days cut-off point, we defined 'episodes' as the group of visits (at least two) separated by $\leqslant 365$ days among them (see Supplemental Figure 1).

The data were obtained during routine clinical practice with the informed consent of patients to be treated in a service that performs clinical care and research, hence there is not written informed consent. HCSC ethics review board approval was obtained as a retrospective study and waiver of informed consent was obtained for the use of de-identified clinical records. Furthermore, the study was conducted according to the Declaration of Helsinki.
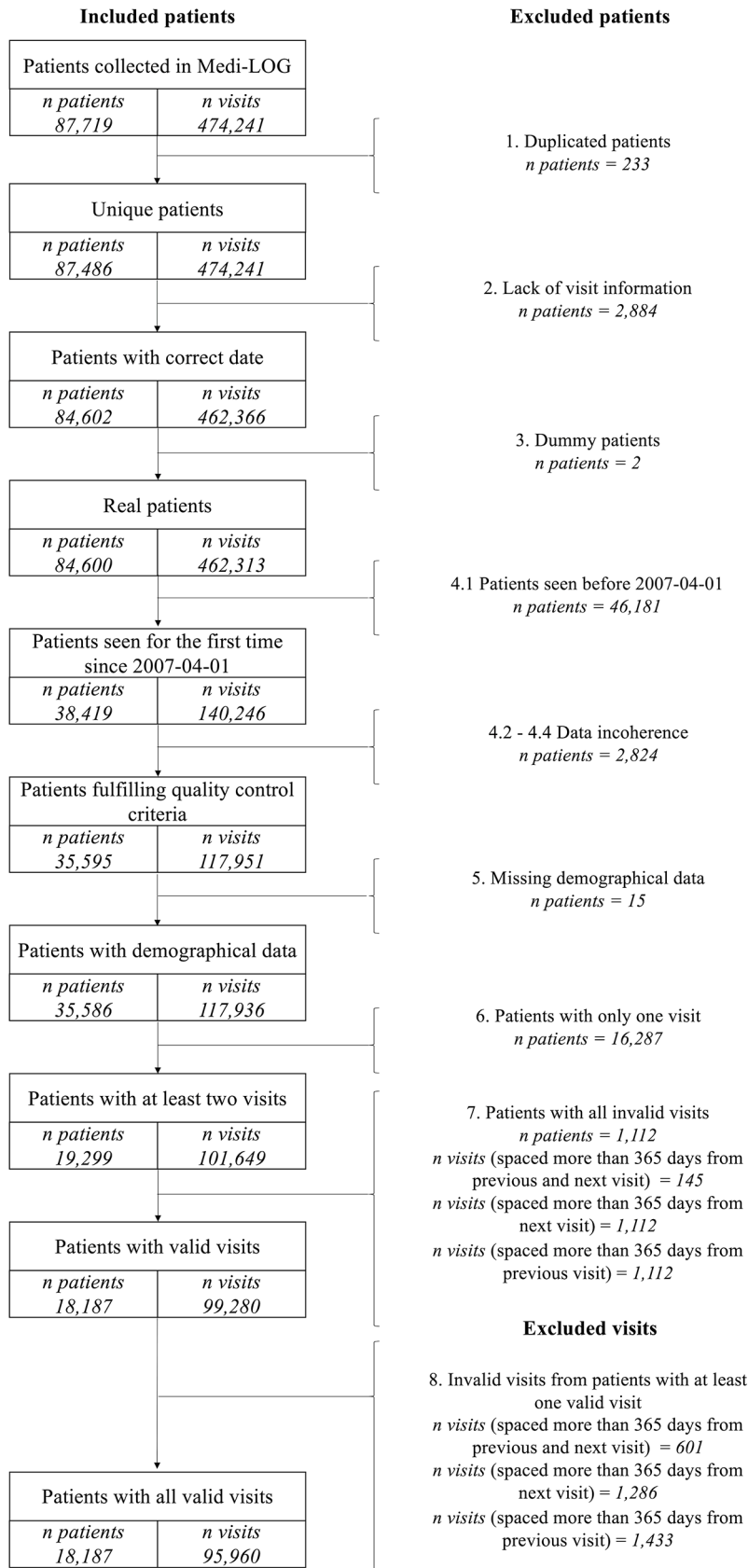
Included patients   **Excluded patients**

Patients collected in Medi-LOG

n patients	*n visits*
87,719	474,241

1. Duplicated patients
n patients = 233

Unique patients

n patients	*n visits*
87,486	474,241

2. Lack of visit information
n patients = 2,884

Patients with correct date

n patients	*n visits*
84,602	462,366

3. Dummy patients
n patients = 2

Real patients

n patients	*n visits*
84,600	462,313

4.1 Patients seen before 2007-04-01
n patients = 46,181

Patients seen for the first time
since 2007-04-01

n patients	*n visits*
38,419	140,246

4.2 - 4.4 Data incoherence
n patients = 2,824

Patients fulfilling quality control
criteria

n patients	*n visits*
35,595	117,951

5. Missing demographical data
n patients = 15

Patients with demographical data

n patients	*n visits*
35,586	117,936

6. Patients with only one visit
n patients = 16,287

Patients with at least two visits

n patients	*n visits*
19,299	101,649

7. Patients with all invalid visits
n patients = 1,112
n visits (spaced more than 365 days from
previous and next visit) *= 145*
n visits (spaced more than 365 days from
next visit) *= 1,112*
n visits (spaced more than 365 days from
previous visit) *= 1,112*

Patients with valid visits

n patients	*n visits*
18,187	99,280

Excluded visits

8. Invalid visits from patients with at least
one valid visit
n visits (spaced more than 365 days from
previous and next visit) *= 601*
n visits (spaced more than 365 days from
next visit) *= 1,286*
n visits (spaced more than 365 days from
previous visit) *= 1,433*

Patients with all valid visits

n patients	*n visits*
18,187	95,960

Figure 1. Flow chart of patients and visits included in the study after applying the inclusion and exclusion criteria.

Variables

Our primary outcome was the HRQoL, measured using the RCI,[27] and assessed in every patient's visit. The RCI comprises two components, disability, and the distress levels (see Supplemental Table 1). The former covers the daily life activities of individuals (i.e. general mobility, self-care, usual activities, and social relationships) by considering eight ordinal levels ranging from no disability to unconsciousness. The latter is measured as emotional (i.e. mood) or physical (i.e. pain) distress and is defined as the worst measure between them. Distress has four different categories ranging from no distress/no pain to a high level of distress/pain. Up to 29 Rosser health states arouse by combining these two components. The key advantage of the RCI is its simplicity; in fact, a patient can be classified in less than 30 seconds, which overcomes the time-consumption drawback from other quality of life (QoL) measures.

As predictors, we included demographic and clinically related variables (diagnoses, treatments, comorbidities, and concomitant treatments) collected during the patients' visits using our EHR.

Comorbidities and concomitant treatments included not only RMD-related variables, but also non-RMD variables. For instance, chronic kidney disease, neurological disease, asthma, anemia, psychiatric conditions, or hyperlipidemia are some examples of the comorbidities studied. Moreover, drugs such as insulin, tramadol, thyroid hormone, or surgery-related predictors such as thyroidectomy, appendicitis or articular prosthesis were also considered.

After a feature selection process (see Supplemental data 'Predictors'), the number of categories related to diagnoses was 34, the number of categories related to treatments was 33, the number of comorbidities was 87, and the number of concomitant treatments was 34.

Based on our previous definition of episodes, we also considered 'diagnoses and treatment episodes'. They were defined as consecutive follow-up periods (i.e. the end of one episode matches with the beginning of the next episode) for a particular diagnosis or treatment. These episodes begin in the visit when the patient receives that diagnosis or is prescribed with that treatment, and: (a) in the previous visit the patient had not received such diagnosis or treatment: in this case, the previous episode (if there was one) ends in the same visit as the new episode begins, or (b) the previous visit took place more than 365 days apart: in this case, the previous episode (if there was one) ends in that previous visit more than 365 days apart.

Based on the beginning and end of each episode, the duration from the first visit of the episode until each visit of the episode was calculated. In addition, the cumulative duration was calculated considering the total time that a patient has received a treatment or a diagnosis from the very first visit until each visit. For further details see Supplemental Table 2.

Episodes for each diagnosis and treatment were categorized based on their chronological order (1st episode, 2nd episode, 3rd episode, and so on). Furthermore, episodes with a reduced number of visits (<1% of 95,960) were grouped with those episodes taking place earlier to comply with the 1% criteria (see Supplemental Excel, DxT episodes).

For each visit, the patient's RCI mean and median values of the previous visits in a period of 91, 182, 274, and 365 days were computed. If, for a particular visit, the patient had not attended our clinic in the time periods considered (or in the case of the first visit to our clinic); we assumed that their previous averaged HRQoL was the same as in the actual visit (we referred to this process as 'imputation').

A more detailed explanation can be found in Supplemental data 'Quality of Life Imputation' section and Supplemental Table 3.

Patients' occupation or demographical variables such as age at the first visit or the total number of visits per patient were also considered.

Overall, 410 different variables were included in the analyses (see Supplemental data 'Predictors' section, Supplemental Excel file predictors and Supplemental Excel file categories: 'Final diagnoses' and 'Final treatments' sheet).

Statistical analyses

The association between demographic and clinical variables and repeated measures of HRQoL was analyzed using generalized estimating equations (GEEs)[28] and penalized generalized

estimating equations (PGEEs). Three different working correlation structures were used in both the GEE and PGEE models: independence, exchangeable, and first-order autoregressive model AR(1). The variable selection process consisted of four sequential steps: GEE bivariable analyses, removal of highly correlated predictors, multivariate GEE analyses and PGEE analyses. For further details see Supplemental data 'Statistical analyses'.

These steps were repeated for independence, exchangeable and AR(1) correlation structures. Based on the root mean square error (RMSE) values of the three different penalized models, the correlation structure with the lowest score was chosen.

Statistical analyses were performed using R statistical software version 3.3.2[29] and the 'PGEE' package.[30]

Model explanation and interpretability

Clustering techniques were used to validate the results from the PGEE analyses, depending on the final selected predictors' type, such as k-means in case all the predictors were continuous variables;[31] or the Kamila semi-parametric algorithm, in case a combination of continuous and categorical predictors were selected.[32] A summary of the variables selection and validation process can be found in Supplemental Figure 2.

Results

Patients' description

After applying the selection criteria (see Figure 1), 18,187 patients with 95,960 observations (visits or contacts) were included. Table 1 and Supplemental Table 4 show the patients' demographic and clinically related characteristics at the first valid visit.

There were 16,139 (88.74%), 1825 (10.03%), 205 (1.13%) and 18 (0.1%) patients with up to one, two, three and four different follow-up episodes (i.e. group of visits spaced more than 365 days). From the 95,960 visits, 87,097 (90.76%), 8005 (8.34%), 803 (0.84%) and 55 (0.06%) belonged to the first, second, third and fourth follow-up episode, respectively. The percentage of imputed median and mean RCI values (excluding first visits) were 29.3% (91 day

period), 10.7% (182 days), 4.7% (274 days) and 2.1% (365 days). All first visits (18.9%) were imputed due to the lack of previous health status values (see Supplemental Table 3).

The most prevalent diagnosis in the first valid visit is 'no diagnoses'. This code is given when the precise diagnosis and/or the disease etiology is not clear and further tests such as laboratory or imaging are necessary to provide an accurate diagnosis. Considering all visits, 3594 patients (19.8%) received the 'no diagnoses' code at least once during their follow-up. Of them, 3115 patients (86.7%) receive this diagnosis in the first valid visit to our clinic. Furthermore, the 'no diagnoses' code is modified to a RMD diagnosis code in the next visit in 1560 patients (50.1%), and over the total course of the follow-up, in 2033 patients (65.3%). The most frequent RMD diagnoses given after a 'no diagnoses' code was back pain (9.5%), joint pain (8.0%), tendinitis (upper extremities: 7.8%), tendinitis (lower extremities: 6.4%), peripheral osteoarthritis (4.5%), knee osteoarthritis (4.5%), chronic polyarthritis (4.4%), peripheral neuropathy (4.1%), and rheumatoid arthritis (3.9%).

Model development

We present the results for the independent correlation structure, as it showed the lowest RMSE value (3.883, 4.033, and 3.902 for the independent, exchangeable, and AR(1) structures, respectively).

After carrying out the bivariate analyses, 292 predictors with a *p*-value < 0.15 were carried out to the next step. Afterwards, 49 predictors were removed due to their high pairwise correlation. From the remaining 243 predictors, 214 were combined into sets of clinically related terms ($n = 66$ different sets). After subsequent multivariate analyses, 125 predictors remained, and those without statistical significance were removed ($n = 89$). Predictors that could not be included in any set of clinically related terms were incorporated directly in the PGEE model ($n = 28$). Therefore, 153 predictors were included in the PGEE analysis; two of them were related to the patients' averaged previous HRQoL before each particular visit. A full description of the removed predictors along these steps can be found in Figure 2 and in the Supplemental Excel file 'Predictors selection independent'. In addition, in the Supplemental Excel files 'Predictors selection

Table 1. Demographic and clinically related patient characteristics at the first valid visit, including the five most prevalent diagnoses and treatments given and prescribed by the rheumatologists, and comorbidities and concomitant treatments given by other physicians.

Variables	Number of patients (n = 18,187)
Women, n (%)	12,623 (69.41%)
Age, median (Q1–Q3)	56.39 (44.45–70.78)
Rosser index, median (Q1–Q3)	98.6 (97.2–99.5)
Pain/distress level, n (%)	
None	2547 (14.00)
Low	11,345 (62.38)
Moderate	4150 (22.82)
High	145 (0.8)
Disability level, n (%)	
None	5898 (32.43)
Slight social	7565 (41.60)
Severe social and slight physical	2693 (14.81)
Moderate decrease in mobility	1336 (7.35)
Severe decrease in mobility	531 (2.92)
Almost dependent	160 (0.88)
In bed	4 (0.02)
Unconscious	0 (0)
Occupation, n (%)	
Active	10,115 (55.62)
Housework	4587 (25.22)
Retired	3096 (17.02)
Student	389 (2.14)
Diagnoses, n (%)	
No diagnoses	3115 (17.13)
Tendinitis (upper extremities)	2590 (14.24)

(Continued)

Table 1. (Continued)

Variables	Number of patients (n = 18,187)
Back pain	1808 (9.94)
Pain in joint	1773 (9.75)
Osteoarthritis of knee	1220 (6.71)
Number of diagnoses, median (Q1–Q3)	1 (1–1)
Treatments, n (%)	
Analgesic 1st level	5236 (28.79)
NSAIDs	5111 (28.1)
Gastric protector	2984 (16.41)
Corticosteroid	2340 (12.87)
Calcium and vitamin D	1639 (9.01)
Number of treatments, median (Q1–Q3)	1 (0–2)
Comorbidities and concomitant treatments, n (%)	
Hypertension	4432 (24.37)
Dyslipidemia	3733 (20.53)
Diabetes mellitus	1461 (8.03)
Hypothyroidism	1114 (6.13)
Cancer	872 (4.79)
Number of comorbidities, median (Q1–Q3)	2 (0–5)

NSAIDs, non-steroidal anti-inflammatory drugs; Q1–Q3, first and third quartiles.

exchangeable' and 'Predictors selection AR(1)', the same process for the other two working correlation structures can be found.

Two predictors related to the averaged HRQoL values from previous periods remained in the fourth step for the independent working correlation structure (mean RCI value over the last 91 days and median RCI value over the last year). Therefore, two grid searches for estimating the λ penalty tuning parameter were performed: one for the model that included the averaged previous HRQoL predictors and one for the model without those variables. Up to 29 different values were tested under k-fold =5;
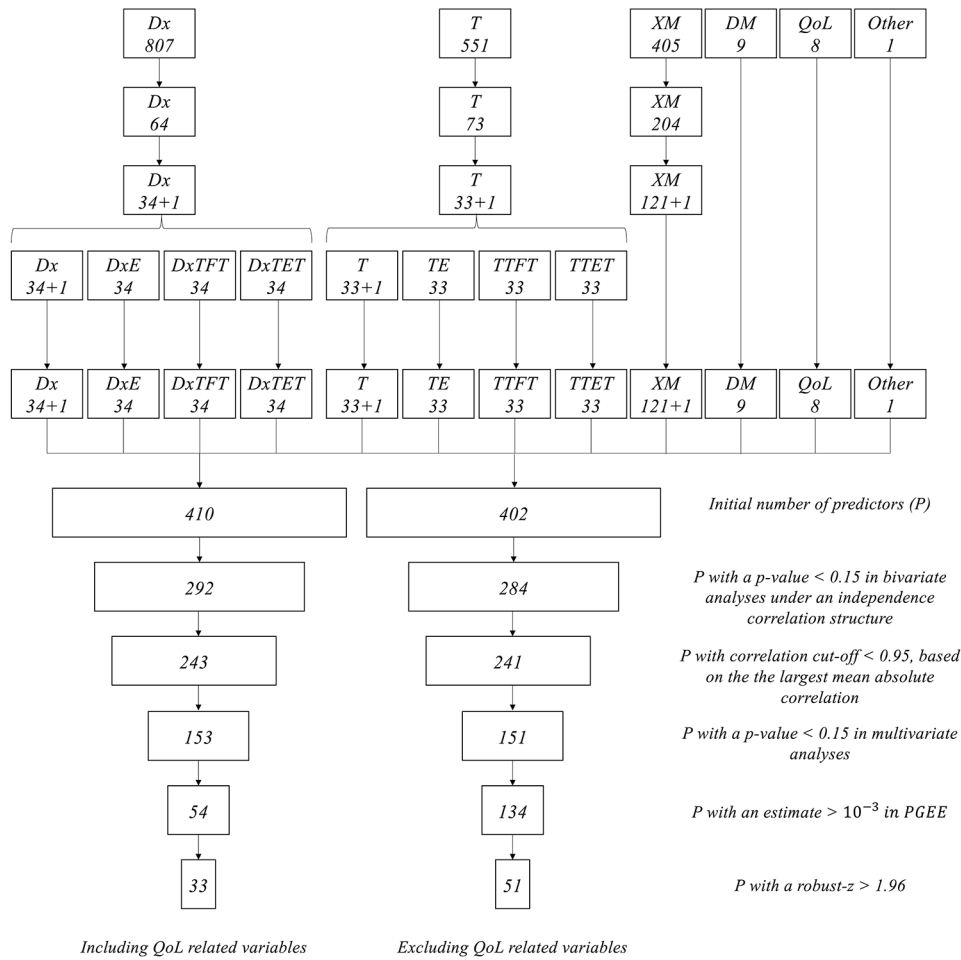
Figure 2. Flow chart of the number of predictors during the selection process.
DM, demographic variable; Dx, diagnosis-related variable; DxE, diagnosis episode-related variable; DxTFT, diagnosis total follow-up time; DxTET, diagnosis episode follow-up time; QoL, quality of life-related variable; T, treatment variable; TE, treatment episode-related variable; XM, comorbidity variable.

the minimum tested penalty value was 0.001, whereas the maximum penalty value was 1. The rest of the PGEE tuning parameters, such as the working correlation structure, the link and variance functions were defined as independence, Gaussian and identity, respectively. Additional details of the tested penalty values for the working correlation structures with and without averaged previous HRQoL predictors can be found in Supplemental Excel file 'Penalty tuning parameter' and in Supplemental Figures 3 and 4.

Selected predictors

The λ that yielded the smallest prediction error was λ = 0.04 for the model that included the averaged previous HRQoL predictors and 0.004 for the model without them.

For the first model, 53 predictors were selected after running the PGEE model and only 33 of them had statistical significance (*p*-value < 0.05). Table 2 and Supplemental Figures 5 and 6 show the demographic, diagnoses, treatments, comorbidities, and averaged previous HRQoL predictors negatively and positively associated with HRQoL.

For the model that did not include HRQoL-related predictors, 51 significant variables were finally selected. For this model, predictors negatively and positively associated with HRQoL can be found in Table 3 and Supplemental Figures 7 and 8.

Table 4 shows the predictors shared between the two models. We observed that the effect's

Table 2. Selected predictors after applying independent working correlation structure.

Predictor	Episode	Estimate	*p*-Value
Sociodemographic			
(DM) Sex		−0.077	0.013
Diagnoses			
(Dx) Fibromyalgia		−0.568	<0.001
(Dx) Psoriatic arthritis		−0.26	0.029
(Dx) Muscle disorders		−0.251	<0.001
(Dx) Osteoporosis		0.183	0.003
Diagnoses episode			
(DxE) Monoarthritis	2	−0.634	0.018
(DxE) Other inflammatory diseases CT	1	−0.623	0.033
(DxE) Osteoarthritis	2	−0.235	0.02
(DxE) Raynaud syndrome	1	0.175	<0.001
(DxE) Muscle disorders	1	0.202	<0.001
Quality of life-related			
(HS) median 365		0.172	0.013
(HS) mean 91		0.618	<0.001
Other			
(Other) Occupation (housework)		−0.167	0.004
Treatment			
(T) Anti-gout		−0.4	<0.001
(T) Analgesic 1st level		0.022	0.023
(T) NSAIDs		0.082	<0.001
(T) Corticosteroid		0.184	<0.001
(T) Colchicine		0.27	0.002

(Continued)

Table 2. (Continued)

Predictor	Episode	Estimate	*p*-Value
(T) Analgesic 3rd level		0.412	0.001
Treatment episode			
(TE) Analgesic 3rd level	1	−0.843	<0.001
(TE) DMARDs: azathioprine	1	−0.405	0.011
(TE) Ferrum	1	−0.242	0.032
(TE) Analgesic 1st level	1	0.022	0.034
(TE) Anti-gout	1	0.581	<0.001
Comorbidity (disease)			
(XM) Chronic kidney disease		−0.854	0.01
(XM) Ischemic heart disease		−0.4	0.01
(XM) Neurological disease		−0.29	0.045
(XM) Other virus infection		−0.269	0.029
(XM) Other arrhythmias		−0.247	0.031
(XM) Back pain		−0.161	0.044
Comorbidity (concomitant treatment)			
(XM) Antidepressant medication		−0.016	0.013
(XM) Benzodiazepine		−0.012	0.032
Comorbidity (others)			
(XM) Number of comorbidities given in current visit		0.018	0.009

DM, demographic variable; DMARDs, disease-modifying anti-rheumatic drugs; Dx, diagnosis-related variable; DxE, diagnosis episode-related variable; HS, quality of life-related variable; NSAIDs, non-steroidal anti-inflammatory drugs; T, treatment variable; TE, treatment episode-related variable; XM, comorbidity variable.

Table 3. Selected predictors after applying independent working correlation structure, without considering imputed HRQoL predictors (i.e. averaged and median RCI values).

Predictor	Episode	Estimate	*p*-Value
Sociodemographic			
(DM) Sex		−0.3	0.001
(DM) Baseline age		−0.024	<0.001
Diagnoses			
(Dx) Fibromyalgia		−1.613	<0.001
(Dx) Monoarthritis		−0.901	0.009
(Dx) Axial neuropathy		−0.843	0.005
(Dx) Muscle disorders		−0.48	<0.001
(Dx) Tendinitis		0.147	0.05
(Dx) Osteoarthritis of hand		0.341	0.001
(Dx) Osteoarthritis of first carpometacarpal joint		0.67	<0.001
(Dx) Sjögren syndrome		0.775	<0.001
(Dx) Osteoporosis		0.814	<0.001
Diagnoses episode			
(DxE) Monoarthritis	2	−1.659	<0.001
(DxE) Tendinitis (lower extremities)	1	0.238	0.006
(DxE) No diagnoses	1	0.341	0.001
(DxE) Tendinitis (lower extremities)	2	0.47	0.017
(DxE) Mixed connective tissue disease	1	0.495	0.025
(DxE) Muscle disorders	1	0.598	<0.001
(DxE) Polymyalgia rheumatica	1	0.649	0.002
(DxE) Raynaud syndrome	1	0.851	<0.001

(Continued)

Table 3. (Continued)

Predictor	Episode	Estimate	*p*-Value
Diagnoses total follow-up time			
(DxTFT) Monoarthritis		0.003	<0.001
Other			
(Other) Occupation (housework)		−0.467	<0.001
Treatment			
(T) Number of treatments given in current visit		−0.138	0.014
(T) Calcium and vitamin D		0.459	0.001
(T) Colchicine		0.476	0.002
(T) Statins		0.485	0.01
(T) BIO adalimumab		0.536	0.011
(T) DMARDs subcutaneous methotrexate		0.84	<0.001
(T) SYSADOA		0.863	<0.001
(T) DMARDs oral methotrexate		1.132	<0.001
(T) Analgesic 3rd level		1.896	<0.001
Treatment episode			
(TE) Analgesic 3rd level	1	−3.473	<0.001
(TE) DMARDs oral methotrexate	2	−1.139	0.009
(TE) Ferrum	1	−0.75	0.008
(TE) Analgesic 2nd level	1	−0.688	0.001
(TE) DMARDs oral methotrexate	1	−0.665	0.01
(TE) Antidepressant	1	−0.658	0.01
(TE) NSAIDs	1	0.378	<0.001
(TE) NSAIDs	2	0.319	0.045

(Continued)

Table 3. (Continued)

Predictor	Episode	Estimate	*p*-Value
(TE) Gabapentin	1	0.479	0.038
(TE) Anti-gout	1	0.679	<0.001
Comorbidity (disease)			
(XM) Chronic kidney disease		−2.954	0.017
(XM) Neurological disease		−1.931	<0.001
(XM) Fall		−1.456	0.015
(XM) Diabetes mellitus		−1.169	0.004
(XM) Ischemic heart disease		−1.152	0.027
Comorbidity (concomitant treatment)			
(XM) Antidepressant medication		−0.912	<0.001
(XM) Other virus infection		−0.821	0.017
(XM) Inhaled corticosteroid		−0.63	0.028
(XM) Proton pump inhibitor medication		−0.408	0.011
(XM) Angiotensin II receptor blocker		0.346	0.01
(XM) Benign breast tumor		0.364	0.028
(XM) ORL surgeries		0.46	0.001
(XM) Oral antidiabetics		0.828	0.025
Comorbidity (others)			
(XM) Number of comorbidities given in current visit		0.147	<0.001

DM, demographic variable; DMARDs, disease-modifying anti-rheumatic drugs; Dx, diagnosis-related variable; DxE, diagnosis episode-related variable; DxTFT, diagnosis total follow-up time; HS, quality of life-related variable; NSAIDs, non-steroidal anti-inflammatory drugs; ORL, otorhinolaryngology; T, treatment variable; TE, treatment episode-related variable; XM, comorbidity variable.

direction was the same for both models, although its magnitude was lower when including the averaged previous HRQoL predictors. CKD, a first episode of third level analgesic treatment, a second episode of monoarthritis, and the diagnosis of fibromyalgia were the predictors with the negative effect on the HRQoL. Conversely, a first episode of anti-gout (or uric acid-lowering) medication and the prescription of third level analgesic treatment or colchicine in the previous visit showed the greatest positive effect on the HRQoL.

Model validation
To validate our results, we performed clustering techniques using the PGEE models' selected predictors as they were supposed to be closely related to the patients' HRQoL. Clusters of patients with different mean RCI values were expected to be found. Because both models included continuous (such as the number of comorbidities) and categorical data (such as the presence or absence of a prescription or a diagnosis), the Kamila algorithm was employed for clustering. Clusters with four to six groups were calculated for both models.

For the model including averaged previous HRQoL predictors, observations from clusters were separated by their HRQoL, regardless of the number of groups. Conversely, for the model that did not include the HRQoL-related variables, observations were not split by the RCI value. A summary of the RCI median values by cluster in both models, depending on the number of clusters, can be found in Supplemental Table 5.

Discussion
We have studied the effect of demographic and clinically related variables in the HRQoL of patients with RMDs attending a rheumatology outpatient clinic. Models were built using PGEE, based on a set of predictors extracted from an EHR, which stores longitudinal patient data over up to 10 years.

After variable selection, 19 variables, including RMDs diagnoses, comorbidities, and treatments, showed a significant association with HRQoL, regardless of the inclusion of predictors related to the patient averaged HRQoL from previous visits.

The variable with the greatest negative influence on the HRQoL of patients affected with RMDs

Table 4. Shared predictors among the models with independent correlation structure.

Predictors	Episode	Including QoL-related variables		Excluding QoL-related variables	
		Estimate	**p-Value**	**Estimate**	**p-Value**
Sociodemographic					
(DM) Sex		−0.077	0.013	−0.3	0.001
Diagnoses					
(Dx) Fibromyalgia		−0.568	0	−1.613	<0.001
(Dx) Muscle disorders		−0.251	0	−0.48	<0.001
(Dx) Osteoporosis		0.183	0.003	0.814	<0.001
Diagnoses episode					<0.001
(DxE) Monoarthritis	2	−0.634	0.018	−1.659	<0.001
(DxE) Raynaud syndrome	1	0.175	0	0.851	<0.001
(DxE) Muscle disorders	1	0.202	0	0.598	<0.001
Other					<0.001
(Other) Occupation (housework)		−0.167	0.004	−0.467	<0.001
Treatment					
(T) Colchicine		0.27	0.002	0.476	0.002
(T) Analgesic 3rd level		0.412	0.001	1.896	<0.001
Treatment episode					<0.001
(TE) Analgesic 3rd level	1	−0.843	0	−3.473	<0.001
(TE) Ferrum	1	−0.242	0.032	−0.75	0.008
(TE) Anti-gout	1	0.581	0	0.679	<0.001
Comorbidity (disease)					
(XM) Chronic kidney disease		−0.854	0.01	−2.954	0.017
(XM) Ischemic heart disease		−0.4	0.01	−1.152	0.027
(XM) Neurological disease		−0.29	0.045	−1.931	<0.001
(XM) Other virus infection		−0.269	0.029	−0.821	0.017
Comorbidity (concomitant treatment)					
(XM) Antidepressant medication		−0.016	0.013	−0.912	<0.001
Comorbidity (other)					
(XM) Number of comorbidities given in current visit		0.018	0.009	0.147	<0.001

DM, demographic variable; Dx, diagnosis-related variable; DxE, diagnosis episode-related variable; T, treatment variable; TE, treatment episode-related variable; XM, comorbidity variable.

was the presence of CKD. This condition has previously been associated with lower HRQoL when compared with the general population.[33,34] In our cohort, the presence of this comorbidity was significantly associated with the diagnoses of gout (17.4% *versus* 3.3% of patients, for those with and without CKD, respectively; *p*-value from χ^2 test $=8.4\times10^{-40}$), which has previously been shown,[35] but also with the diagnoses of crystal arthropathies (8.2% *versus* 2.1%; $p=8.0\times10^{-13}$), rheumatic polymyalgia (7.9% *versus* 3.6%; $p=6.9\times10^{-5}$), and knee osteoarthritis (15.5% *versus* 1.4%; $p=0.03$), among others (Supplemental Excel file CKD). Our results suggest that we should keep in mind the impact of this comorbidity in the care of patients with RMDs, as it could influence their HRQoL regardless of their diagnosis.

As for RMD diagnoses, monoarthritis had the greatest deleterious effect on HRQoL. To the best of our knowledge, no previous studies have analyzed the effect of this diagnosis on HRQoL. At least in our case, monoarthritis seems to be a temporary diagnosis, given that the etiology of this clinical presentation is still not clear: on the one hand, the proportion of visits in which this diagnosis was given diminished during follow-up (from 100% of first episods visits, to 62.9% in the second visits, 46.0% in the third, 37.4% in the fourth, 32.7% in the fifth, and 12.8% in the sixth and subsequent visits). Besides, only in 296 of the 1199 episodes in which monoarthritis was diagnosed (24.7%), no other RMD diagnosis was given. In the other episodes, other RMDs were also diagnosed, such as knee osteoarthritis (322 episodes, 26.9%), gout and crystal arthropathies (241 episodes, 20.1%), and lower limb tendinitis (88, 7.3%). However, we also observed differences when comparing the associated diagnoses given in the first *versus* the second and subsequent monoarthritis episodes: there was a decrease in the proportion of episodes when monoarthritis was the only diagnosis given (29.4% *versus* 7.1%), and an increase in the proportion of other diagnoses, such as knee osteoarthritis (24.4% *versus* 36.1%), or gout/crystal arthropathies (18.2% *versus* 27.4%). These changes in the patterns of diagnosis may contribute to the different HRQoL associated with monoarthritis episodes (Supplemental Excel file Monoarthritis).

On the other hand, fibromyalgia (see Table 2) has previously been associated with lower HRQoL compared with the general population,[36,37] but also when compared with other RMDs, such as rheumatoid arthritis or spondyloarthritis.[38]

Regarding treatments prescribed by the rheumatologists, we observed that uric acid-lowering medication was associated with a significant increase in HRQoL, once the patient received the first prescription (all subsequent episodes were combined due to the rule of 1%). We observed that in the 486 episodes of these medications, the mean (standard deviation) percentage of visits per episode in which uric acid-lowering drugs had been prescribed was 92.6% (18.7). Moreover, we observed that the higher the percentage the better the HRQoL in individual visits, even after adjusting for gender, age, diagnosis of gout in each visit, and prescription of uric acid-lowering drugs in the previous visit [coefficient 0.20 per 10% increase in the percentage of visits per episode in which uric acid-lowering drugs had been prescribed, 95% confidence interval (0.1–0.31), $p=2.0\times10^{-4}$].

We have identified the following study limitations:

1. The misspecification of the working correlation structure can be an issue that could lead to a loss of efficiency in the estimation of the regression parameters.[39] Several approaches to optimize its selection have been developed.[40–44] However, some authors suggest basing the working correlation structure decision on the mechanism in which the data are generated.[45] In our case, we decided to use the RMSE to base the correlation structure selection, meaning that we selected the fittest model for our data.
2. Regarding variable selection, we are aware that a bivariable selection step may introduce bias due to the lack of confounding control, which may result in the rejection or inclusion of inappropriate variables as studied by Sun et al.[46] In our study and due to the large number of predictors tested, we decided to increase the p-value threshold for deleting variables.
3. All predictors categorized as 'treatments' have a pharmacological component. Despite analyzing medical and surgical procedures (as comorbidities), we missed the inclusion of variables that measure the effect of non-pharmacological treatments, such as, physiotherapy[47] or occupational therapy,[48] which have been demonstrated to be associated with the HRQoL of RMD

patients. The lack of these variables is explained by two factors: all patients received disease-specific exercise guidelines by their rheumatologist physician; and these variables are not codified in our EHR. However, non-pharmacological treatment data collected (e.g. assistive technology use) from a multidisciplinary team would potentially enrich the study.

4. Other disability-related predictors such as the patients' degree of disability recognized by current legislation, and potential resources received according to that disability degree would be interesting to consider in future iterations.

5. Respecting the validation of our results using cluster analysis, only when including the variables related to the previous patient's averaged HRQoL, could we discriminate groups of observations with different mean RCI values. These variables may introduce biases in the results as past values were imputed, as described in the Methods section: for the first patient's visit since no previous health status information was available; and for those visits during follow-up with no previous data in the selected periods (91, 182, 274, and 365 days). Despite this issue, we were able to identify 19 predictors shared between the two models, see Table 4, that seem to influence the patients' HRQoL, regardless of the presence or absence of averaged past RCI values.

Concerning the study strengths, we want to point out that GEE has been widely used in different fields, including rheumatology,[49] as these models allow us to take into account both the inter-subject and intra-subject variability, its robustness when the primary outcome variable (RCI) does not have a Gaussian distribution and when varying the working correlation structure, or its ability to use all of the available information even for unequal numbers of repeated measurements and unequal time intervals.[50] Besides, we have used a recent implementation of PGEE in R,[30,51] addressing the analysis of longitudinal data with high-dimensional covariates.[52] This method can perform simultaneous estimation and variable selection by specifying a tuning penalty parameter, λ, which regulates the model complexity by shrinking the coefficients' estimates to zero. A few studies have recently employed this package to study the working correlation structure selection[53] and the separation[54] issues in GEE.

In this study we tried to assess predictors governing the HRQoL of patients who attended an outpatient rheumatology setting, instead of finding predictors associated with concrete RMDs. In this scenario, in which the rheumatic population may be highly heterogeneous, we found that predictors which have a major impact on HRQoL are not diagnoses themselves – excepting fibromyalgia – but treatment requirement and comorbidities predictors. To the best of our knowledge this is the first study attempting to describe – in an outpatient setting – factors affecting the HRQoL of RMD patients. We believe that these results could be helpful for other rheumatic consultations because they reveal novelty factors that may determine the wellbeing of the patient.

Conclusion

In conclusion, we have shown several variables affecting the HRQoL of patients with RMDs attending a tertiary rheumatology outpatient clinic. The predictors with the greatest negative influence were the presence of CKD, monoarthritis, and fibromyalgia. Conversely, the use of uric acid-lowering drugs, third level analgesics and colchicine was associated with the greatest positive influence on HRQoL.

Authors' contributions

LRR, LAA, and EM conceived and designed the study. LLM, EP, JAJ and BFG collected data. AMG, LRR, and EM performed the data analysis and interpreted the data. All authors were involved in the drafting and/or revising of the manuscript. All authors gave final approval of the version of the article to be published.

Conflict of interest statement

The authors declare that there is no conflict of interest.

Patient and public involvement

Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication

Not required.

Ethics approval

The study was approved by the Hospital Clínico San Carlos Ethics Committee (approval number 20/268-E-BS). This study was conducted according to the principles of the Declaration of Helsinki.

ORCID iDs

Alfredo Madrid-García https://orcid.org/0000-0002-1591-0467

Benjamín Fernández-Gutiérrez https://orcid.org/0000-0002-6126-8786

Luis Rodríguez-Rodríguez https://orcid.org/0000-0002-2869-7861

Data availability statement

Data are available on reasonable request.

Supplemental material

Supplemental material for this article is available online.

References

1. Testa MA and Simonson DC. Assessment of quality-of-life outcomes. *N Engl J Med* 1996; 334: 835–840.

2. Guyatt GH, Feeny DH and Patrick DL. Measuring health-related quality of life. *Ann Intern Med* 1993; 118: 622–629.

3. Picavet HSJ and Hoeymans N. Health related quality of life in multiple musculoskeletal diseases: SF-36 and EQ-5D in the DMC3 study. *Ann Rheum Dis* 2004; 63: 723–729.

4. Lin X-J, Lin I-M and Fan S-Y. Methodological issues in measuring health-related quality of life. *Tzu Chi Med J* 2013; 25: 8–12.

5. James SL, Abate D, Abate KH, *et al*. Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet* 2018; 392: 1789–1858.

6. Loza E, Abásolo L, Jover JA, *et al*. Burden of disease across chronic diseases: a health survey that measured prevalence, function, and quality of life. *J Rheumatol* 2008; 35: 159–165.

7. Roux CH. Impact of musculoskeletal disorders on quality of life: an inception cohort study. *Ann Rheum Dis* 2005; 64: 606–611.

8. Rubin R. EULAR: Countries unite to fight rheumatic diseases. *Lancet* 2017; 389: 2276.

9. Canning J, Siebert S, Jani BD, *et al*. Examining the relationship between rheumatoid arthritis, multimorbidity and adverse health-related outcomes: a systematic review. *Arthritis Care Res (Hoboken)*. Epub ahead of print 1 March 2021. DOI: 10.1002/acr.24587

10. Tyack Z, Frakes K, Barnett A, *et al*. Predictors of health-related quality of life in people with a complex chronic disease including multimorbidity: a longitudinal cohort study. *Qual Life Res* 2016; 25: 2579–2592.

11. Beaudart C, Biver E, Bruyère O, *et al*. Quality of life assessment in musculo-skeletal health. *Aging Clin Exp Res* 2018; 30: 413–418.

12. Salaffi F, De Angelis R, Stancati A, *et al*. Health-related quality of life in multiple musculoskeletal conditions: a cross-sectional population based epidemiological study. II. The MAPPING study. *Clin Exp Rheumatol* 2005; 23: 829–839.

13. Ware JE and Sherbourne CD. The MOS 36-item short-form health survey (SF-36). *Med Care* 1992; 30: 473–483.

14. Herdman M, Gudex C, Lloyd A, *et al*. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Qual Life Res* 2011; 20: 1727–1736.

15. Rosser R and Kind P. A scale of valuations of states of illness: is there a social consensus? *Int J Epidemiol* 1978; 7: 347–358.

16. Sociedad Española de Reumatología. Activity indices, questionnaires and other measurement instruments in Rheumatology. https://www.ser.es/profesionales/que-hacemos/investigacion/herramientas/catalina/ (accessed 10 May 2021).

17. Haraldstad K, Wahl A, Andenæs R, *et al*. A systematic review of quality of life research in medicine and health sciences. *Qual Life Res* 2019; 28: 2641–2650.

18. Nikiphorou E, Santos EJF, Marques A, *et al*. 2021 EULAR recommendations for the implementation of self-management strategies in patients with inflammatory arthritis. *Ann Rheum Dis*. Epub ahead of print 7 May 2021. DOI: 10.1136/annrheumdis-2021-220249

19. Senusi AA, Ola D, Mather J, *et al.* Behçet's syndrome and health-related quality of life: influence of symptoms, lifestyle and employment status. *Clin Exp Rheumatol* 2017; 35 (Suppl. 1): 43–50.

20. Campos RP and Vázquez MIR. Health-related quality of life in women with fibromyalgia: clinical and psychological factors associated. *Clin Rheumatol* 2012; 31: 347–355.

21. Chandratre P, Mallen C, Richardson J, *et al.* Health-related quality of life in gout in primary care: baseline findings from a cohort study. *Semin Arthritis Rheum* 2018; 48: 61–69.

22. Haverman L, Grootenhuis MA, van den Berg JM, *et al.* Predictors of health-related quality of life in children and adolescents with juvenile idiopathic arthritis: results from a web-based survey. *Arthritis Care Res (Hoboken)* 2012; 64: 694–703.

23. Wan SW, He H-G, Mak A, *et al.* Health-related quality of life and its predictors among patients with rheumatoid arthritis. *Appl Nurs Res* 2016; 30: 176–183.

24. Kiani AN, Strand V, Fang H, *et al.* Predictors of self-reported health-related quality of life in systemic lupus erythematosus. *Rheumatology* 2013; 52: 1651–1657.

25. Strickland G, Pauling J, Cavill C, *et al.* Predictors of health-related quality of life and fatigue in systemic sclerosis: evaluation of the EuroQol-5D and FACIT-F assessment tools. *Clin Rheumatol* 2012; 31: 1215–1222.

26. Madrid-García A, Font-Urgelles J, Vega-Barbas M, *et al.* Outpatient readmission in rheumatology: a machine learning predictive model of patient's return to the clinic. *J Clin Med* 2019; 8: 1156.

27. Leon L, Rodriguez-Rodriguez L, Aguilar MD, *et al.* Validation of a quality of life instrument in Spanish patients with rheumatic diseases: the Rosser classification system. *J Clin Rheumatol* 2019; 25: 78–84.

28. Halekoh U, Højsgaard S and Yan J. The R package geepack for generalized estimating equations. *J Stat Softw.* Epub ahead of print 22 December 2006. DOI: 10.18637/jss.v015.i02

29. R Development Core Team. A Language and Environment for Statistical Computing. *R Found Stat Comput* 2020. https://www.R-project.org (accessed 15 December 2020).

30. Inan G and Wang L. PGEE: an R package for analysis of longitudinal data with high-dimensional covariates. *R J* 2017; 9: 393.

31. Huang Z. Extensions to the k-means algorithm for clustering large data sets with categorical values. *Data Min Knowl Discov* 1998; 2: 283–304.

32. Foss A, Markatou M, Ray B, *et al.* A semiparametric method for clustering mixed data. *Mach Learn* 2016; 105: 419–458.

33. Pagels AA, Söderkvist BK, Medin C, *et al.* Health-related quality of life in different stages of chronic kidney disease and at initiation of dialysis treatment. *Health Qual Life Outcomes* 2012; 10: 71.

34. Fraser SD, Barker J, Roderick PJ, *et al.* Health-related quality of life, functional impairment and comorbidity in people with mild-to-moderate chronic kidney disease: a cross-sectional study. *BMJ Open* 2020; 10: e040286.

35. Singh JA and Gaffo A. Gout epidemiology and comorbidities. *Semin Arthritis Rheum* 2020; 50: S11–S16.

36. Lee J-W, Lee K-E, Park D-J, *et al.* Determinants of quality of life in patients with fibromyalgia: a structural equation modeling approach. *PLoS One* 2017; 12: e0171186.

37. Verbunt JA, Pernot DH and Smeets RJ. Disability and quality of life in patients with fibromyalgia. *Health Qual Life Outcomes* 2008; 6: 8.

38. Bucourt E, Martaillé V, Goupille P, *et al.* A comparative study of fibromyalgia, rheumatoid arthritis, spondyloarthritis, and Sjögren's syndrome; impact of the disease on quality of life, psychological adjustment, and use of coping strategies. *Pain Med* 2021; 22: 372–381.

39. Crowder M. On the use of a working correlation matrix in using generalised linear models for repeated measures. *Biometrika* 1995; 82: 407.

40. Inatsu Y and Imori S. Model selection criterion based on the prediction mean squared error in generalized estimating equations. *Hiroshima Math J* 2018; 48: 307–334.

41. Cui J and Qian G. Selection of working correlation structure and best model in GEE analyses of longitudinal data. *Commun Stat Simul Comput* 2007; 36: 987–996.

42. Westgate PM. Approaches for the utilization of multiple criteria to select a working correlation structure for use within generalized estimating equations. *Commun Stat – Simul Comput* 2020; 49: 305–316.

43. Fu L, Hao Y and Wang YG. Working correlation structure selection in generalized estimating equations. *Comput Stat* 2018; 33: 983–996.

44. Imori S. Consistent selection of working correlation structure in GEE analysis based on Stein's loss function. *Hiroshima Math J* 2015; 45: 91–107.

45. Breskin A, Cole SR and Hudgens MG. A call for caution in using information criteria to select the working correlation structure in generalized estimating equations. *Epidemiology* 2018; 29: e51.

46. Sun GW, Shook TL and Kay GL. Inappropriate use of bivariable analysis to screen risk factors for use in multivariable analysis. *J Clin Epidemiol* 1996; 49: 907–916.

47. Cuesta-Vargas AI, González-Sánchez M and Casuso-Holgado MJ. Effect on health-related quality of life of a multimodal physiotherapy program in patients with chronic musculoskeletal disorders. *Health Qual Life Outcomes* 2013; 11: 19.

48. Garnaes KK, Mørkved S, Salvesen Ø, *et al.* What factors are associated with health-related quality of life among patients with chronic musculoskeletal pain? A cross-sectional study in primary health care. *BMC Musculoskelet Disord* 2021; 22: 102.

49. Abasolo L, Ivorra-Cortes J, Leon L, *et al.* Contribution of the bone and cartilage/soft tissue components of the joint damage to the level of disability in rheumatoid arthritis patients: a longitudinal study. *Clin Rheumatol* 2019; 38: 691–700.

50. Liang K-Y and Zeger SL. Longitudinal data analysis using generalized linear models. *Biometrika* 1986; 73: 13–22.

51. DaSilva AW, Huckins JF, Wang R, *et al.* Correlates of stress in the college environment uncovered by the application of penalized generalized estimating equations to mobile sensing data. *JMIR Mhealth Uhealth* 2019; 7: e12084.

52. Wang L. GEE analysis of clustered binary data with diverging number of covariates. *Ann Stat* 2011; 39: 389–417.

53. Inan G, Latif MAHM and Preisser J. A PRESS statistic for working correlation structure selection in generalized estimating equations. *J Appl Stat* 2019; 46: 621–637.

54. Mondol MH and Rahman MS. Bias-reduced and separation-proof GEE with small or sparse longitudinal binary data. *Stat Med* 2019; 38: 2544–2560.