Hereditary, socio-behavioural, and immuno-hormonal predictors of incident rheumatoid arthritis and therapy response influences on survival *versus* matched control subjects using a generalised structural equation model

J. Ren[1], A.T. Masi[2], J.C. Aldag[2], C.V. Asche[1,3]

[1]Center for Outcomes Research, Department of Medicine, [2]Department of Medicine, University of Illinois College of Medicine at Peoria, Peoria, IL, USA; [3]Center for Pharmacoepidemiology and Pharmacoeconomic Research, University of Illinois at Chicago College of Pharmacy, Chicago, IL, USA.

Abstract

Objective

Incident onset and survival outcomes involve multiple risk factors and complex interactions preferably investigated in a single study. A generalized structural equation model (GSEM) was used to build an integrative framework to analyse multiple risk factors for incident rheumatoid arthritis (RA) and factors affecting long-term survival outcome.

Methods

Incident RA cases (n=54) had onsets between 1977 and 1994, after cohort entry in 1974. Four cohort control (CN) subjects (n=216) were matched on entry to each case in the community-based CLUE cohort and 270 subjects were followed from 1995 through 2017. Baseline variables included demographic, RA family history, behavioural factors and z-score levels of serum immunological, cytokine, isotype rheumatoid factors (RFs), adrenal steroids, luteinising hormone, prolactin and sex steroids. Four numerical integration methods of GSEM were performed in Stata 15.

Results

Cohort entry factors predicting RA onset included family history of RA, cigarette smoking and IgM RF. Total survival time from cohort entry was associated with incident RA and baseline variables of age, years of completed education, cigarette smoking, immunoreactive proteins and androgenic-anabolic steroids. Mortality of RA was significantly greater than CN subjects for cases having less than good therapy responses in 1995 and only for RA onset before age 60 years. Androgenic-anabolic steroid z-scores significantly correlated with improved survival only in CN subjects with assigned onset before the age of 60.

Conclusion

Successful use of GSEM is feasible in analyses of prospective incident and subsequent survival data and promises to advance understanding of risk factors, survival, and casual pathways.

Key words

Generalized structural equation model, risk factors, rheumatoid arthritis, incident onset, survival outcome

Jinma Ren, PhD
Alfonse T. Masi, MD, DrPH
Jean Aldag, PhD
Carl V. Asche, PhD

Please address correspondence and reprint requests to:
Alfonse T. Masi,
University of Illinois College of Medicine at Peoria,
One Illini Drive,
Peoria, IL 61605, USA.
E-mail: amasi@uic.edu

Received on June 30, 2019; accepted in revised form on August 6, 2019.

© Copyright CLINICAL AND EXPERIMENTAL RHEUMATOLOGY 2020.

Competing interests: none declared.

Introduction

Rheumatoid arthritis (RA) is a common chronic, systemic, inflammatory disease characterised by synovitis and autoantibodies which affects over 1.5 million adults in the United States (1, 2). Its prevalence increases in the elderly; RA affects 0.5 to 1.0 percent of adults and 2 percent of persons 60 years of age or older (3, 4). Due to its systemic nature, RA manifests a variety of articular and extra-articular co-morbidities which contribute to pain, disability, and premature death. Management of RA is complicated, and its frequent prevalence and high cost of care causes this disease to be an area of intense research (3, 5).

While previous epidemiological research has identified risk factors for RA, its aetiology remains undetermined (1, 6). A hypothesis is that interactions between genetics, inflammation, and aging are associated with the onset of RA (6, 7). Mutations in genes encoding the major histocompatibility complex (MHC) are associated with the genomic region encoding pro-inflammatory cytokines (8). Increased serum levels of TNF-α and its downstream inflammatory cytokines, IL-1, IL-6, and IL-8, occur in aging (8). Inflammatory cytokines (9), along with steroid and hormone levels (10), may be potential contributors to immune reactions or risk of RA onset.

We previously identified preclinical biomarkers associated with both incident RA and its mortality (11). Behavioural risk factors, such as years of completed education and smoking, were found to be associated with increased incidence and mortality of RA (3, 4, 12, 13). Conventional statistical methodologies, however, have not been able to effectively demonstrate interrelations among such complicated causal pathways.

Recently, a structural equation model (SEM) has been successfully used to explore risk factors for disorders, such as diabetes mellitus (14) and obesity (15). Generalized SEM (GSEM) application in epidemiology is still new, especially for causal pathway inference. This model has evolved (16, 17) and has recently been developed in statistical software (18), which includes new features for modelling GSEM with survival-time, censored data, multilevel data, latent variables, or other diverse situations. Generalised SEM now offers new, promising methodologic approaches to model and interpret interactions of multiple known risk factors in the disease expression of RA. No report has applied GSEM to investigate the interactions of aging and risk factors for either RA onset or survival to our knowledge.

This study aims to construct an integrative, analytic GSEM framework to investigate risk factors and mediators for incident RA onset and long-term survival as applied to a large prospective, community-based cohort. This model can identify novel causal pathways and reveal potential mediated relations or latent variables linked to incident onset and survival of RA as compared to matched non-RA control (CN) subjects.

Methods

Study design

Data were derived from a 43-year longitudinal, retrospective-prospective database of 54 pre-RA cases and 4:1 matched non-RA CN subjects (n=216), which were nested in the CLUE I cohort (<https://epi.grants.cancer.gov/Consortia/members/clue.html>, Campaign Against Cancer and Stroke) (9-11, 19-24). Control subjects were matched (4:1) to each RA case on race (all Caucasians), sex, and age (within 1 year). The Johns Hopkins University School of Public Health initiated the CLUE I cohort consisting of 12,381 female and 8,680 male residents of Washington Country, Maryland in 1974. Among them, 54 pre-RA and 216 matched CN subjects were identified who had entered into the cohort. Those 270 subjects were assigned into a sub-cohort which constituted the RA Precursors Study (RAPS), which was established at the University of Illinois College of Medicine at Peoria (UICOM-P) in 1991 (9) and continues to the present.

In the GSEM, five hypothesised categories of variables were incorporated in the goal to create an integrated analytic frame to investigate baseline cohort risk factors for incident RA onset and subsequent length of survival: 1. known

risk factors for incident RA – positive family history, cigarette smoking, and positive rheumatoid factor (RF) (3, 12); 2. previously identified baseline immunologic variables associated with RA development in this cohort – elevated serum immune biomarkers, *e.g.* soluble interleukin-2 receptor alpha (sIL-2Rα) (11, 21); 3. previously identified baseline hormonal imbalance associated with increased risk of RA onset, *e.g.* lower androstenedione level (11); 4. hypothesised age-dependent shorter survival time of RA cases than non-RA CN persons (3, 7); and 5. hypothesised interactions between age and hormonal markers affecting survival of both RA and CN subjects (21).

The non-baseline factors of age at RA onset (<60 *vs.* 60+ years) and response to RA therapy (good *vs.* fair or limited) were determined for the cases and assigned to matched CN within sets. In 1995, the incident RA cases were categorised into 3 patterns of course-wide therapy responses by the sole community rheumatologist (good = 1, n=19; fair = 2, n=23; limited = 3, n=12) (21). Therapy response categories reflected the rheumatologist's course-wide evaluation plus dichotomies of swollen joints (<10 *vs.* 10+) and functional capacity (full vs compromised). The good category was assigned to good clinical response plus <10 swollen joints *and* full function; the fair category was assigned to good or fair clinical response plus *either* <10 swollen joints *or* full function; the limited category was assigned to fair or limited clinical response with *neither* <10 swollen joints *nor* full function (21). The non-baseline onset age and therapy response variables were studied separately from the GSEM model as possible influences on RA and CN survival, without multivariable adjustment. Thirty-year and twenty-year survival frequencies were analysed for RA and CN subjects by onset ages <60 years and 60+ years, respectively, and by therapy response categories, and in total subjects.

Study measurements

The European League Against Rheumatism (EULAR) recommendations were used for designation of pre-RA individuals at risk of RA (25). After cohort entry in 1974, all pre-RA cases were diagnosed and confirmed as having the clinical onset of disease according to the American College of Rheumatology (ACR) revised classification criteria (26) in the practice of the sole community rheumatologist (21). Survival years was defined as the interval between the 1974 date of cohort enrollment and mortality or censoring date. Observations were considered censored when patients died in the study period or follow-up ended on December 31, 2017. Demographic variables (age, sex, and years of completed education) and cigarette smoking were collected at baseline entry. Entry age indicates subject's age in years at the time of cohort enrollment. The RA onset age was determined from the rheumatologist's clinical records and dichotomised into <60 and 60+ years to analyse interactions with other variables. Sex was converted into a numeric variable (0-female, 1-male). Race was removed from the analysis because all subjects were Caucasian. Education was coded as the number of years of completed education. The numbers of cigarettes smoked per day at baseline was recorded and grouped into four categories: (1) non-smoker or previous smoker; (2) smoke 1-10; (3) 11-20, and (4) 21+ cigarettes per day. Family history of RA was defined as RA in a first-degree relative at the time of RA case diagnosis (1977-1995), as determined from rheumatologist's records.

Latent variables are unobserved inferred variables that are not measured directly, but may affect outcomes of interest (incident RA and survival time) (27). Since latent variables cannot be measured directly, a group of related variables were used as representation (Fig. 1). Baseline overall latent activity of immunoreactive proteins was derived from C-reactive protein (CRP), soluble interleukin-2 receptor alpha (sIL-2Rα), and soluble tumour necrosis factor alpha-receptor 1 (sTNF-R1). Baseline latent activity of inflammatory cytokines were represented by interleukin-1 beta (IL-1ß) and tumour necrosis factor-alpha (TNF-α). Baseline overall latent level of steroids synthesised by 21-hydroxylase were represented by 17-hydroxyprogesterone and 17-hydroxypregnenolone. Baseline overall latent variable of androgenic-anabolic steroids included dehydroepiandrosterone (DHEA), androstenedione, and sex hormones testosterone, and estradiol (Fig. 1). From 1992 to 1996, the study subjects' stored (-70°C) sera were periodically donated by CLUE I to RAPS for assays of serum acute phase proteins, inflammatory cytokines, RFs, steroidal biomarkers, and hormonal biomarkers, which were performed in national referral laboratories contracted by Dr Masi and UICOM-P (9, 11).

Sample size

Sample size estimation for SEM remains inconsistent in the literature (28-30), although most researchers suggest that at least 15 cases per measured variable are needed. In our study, the total number of 270 subjects should be an acceptable sample size to examine the 16 observed risk factors in our final model.

Data preparation

All steroidal and hormonal (neuroendocrine) values were first transformed to natural logs to improve their symmetry and distributions. Extreme outliers were identified in several variables, particularly the acute phase proteins, which is expected for these sensitive and highly reactive tests. The log-transformed biomarker values were further converted to z-scores to standardise their variances within the 3 separate subject subsets, that is, early-set females, late-set females, and males, as well as within a merged total data set (22). Most individuals had a complete dataset except for a few missing values in immunological and hormonal tests. For missing values, we applied multivariable regression to impute one completed data set for analyses based on the observed data (22, 31). Multiple imputation was not conducted before GSEM because it consumed enormous time to run the model.

Statistical analysis

Based upon our hypothesised categories of risk factors, a schema of GSEM for modelling RA onset and subsequent survival was developed, including all proposed variables, observed and

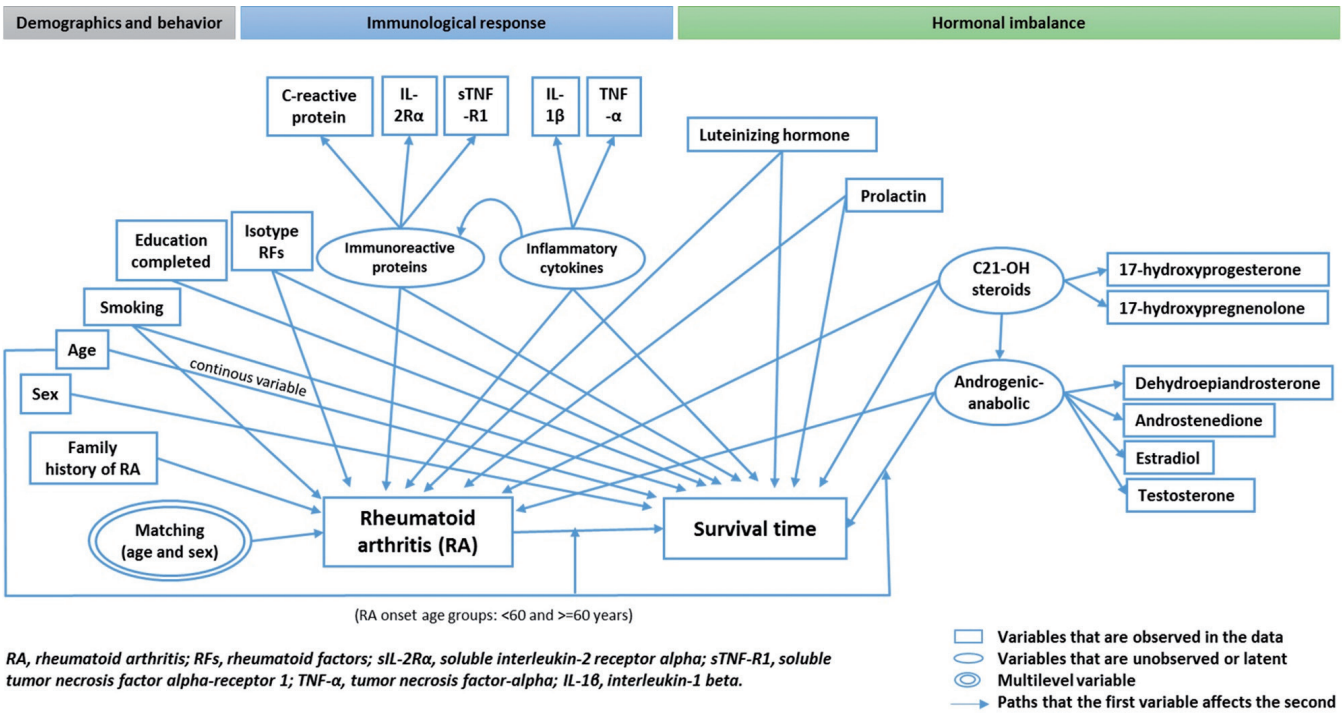

Fig. 1. A proposed structural equation model for incident rheumatoid arthritis onset and survival.

unobserved or latent (Fig. 1). In the causal pathway image, variables were outspread across the group categories: (1) hereditary, demographic, and behaviour; (2) immunological response, (3) hormonal imbalance (Fig. 1). The baseline variables observed in the data (rectangles) were merged into latent variables (ellipses) in the categories of immunological response and hormonal imbalance (Fig. 1).

In the GSEM, the RA variable was stratified into onset age groups of <60 and 60+ years. The factors related to RA onset included smoking, family history of RA, IgM-specific RF, age and sex (matched in the initial cohort), and the biomarkers indicated below. The factors hypothesised to affect mortality (survival) included: age; sex; smoking; years of education completed; immunoreactive proteins (latent variable); inflammatory cytokines (latent variable); LH; prolactin; C21-OH steroids (latent variable); and androgenic-anabolic steroids (latent variable) (Fig. 1). The primary outcome of survival time was defined as the time period between the first RA diagnosis (onset) and death (or censoring time of 12-31-2017) with the same RA values assigned to the matched CN subjects.

All analyses were performed in statistical software Stata 15 (StataCorp LLC). In our proposed GSEM, RA onset and survival were fit by binomial and Weibull distributions, respectively. When modelling RA onset, the matching identifier was set as a multilevel variable. Our model was successfully applied in the non-adaptive Gauss-Hermite quadrature (18). This method helped our model converge within 12 hours on our computer (Windows 7 Ultimate). In the non-GSEM sub-analysis of 30- and 20-year survivals after <60 *vs*. 60+ years onsets, Chi-square and Fisher's exact test were used to test significance of differences of mortality within and between RA and CN groups. A statistical significance level of 0.05 was set for all hypothesis tests. The probability association between each predictor variable and outcome is adjusted for other covariates in the GSEM

Results

Participant characteristics

The total 270 study subjects were composed of 54 entry pre-RA and 216 CN, 180 (66.7%) females (36 pre-RA and 144 CN) and 90 (33.3%) males (18 pre-RA and 72 CN). At cohort entry, the mean (±SD) age was 43 (±11) years and at RA diagnosis, was 55 (±12) years. All individuals were Caucasian. The overall mortality through 2017 was 72.2% for RA cases and 52.8% for CN subjects (p=0.014).

Predictors of incident RA onset

Risk factors for incident RA onset primarily included family history of RA [odds ratio (OR) = 7.2, 95% confidence interval (CI): 2.3–22.4, p=0.001] and IgM rheumatoid factor (RF) [OR= 2.0, CI: 1.4–2.8, p<0.001] (Table I). Cigarette smoking at entry also significantly increased RA onset risk compared to non-smokers, detectable at greater than 10 daily (11–20 /day, OR=2.5, CI: 1.0–6.0, p=0.042 and 21+/day, OR=5.0, CI: 1.8–13.6, p=0.002) (Table I). The 3 preceding risk factors are seen as positive (+) vectors between the observed baseline variables and the incident RA outcome in the GSEM image (Fig. 2) and only smoking decreased survival.

Predictors of mortality

Total mortality was significantly increased with increased cohort entry age (p<0.001). Greater mortality was noted in: (1) RA than non-RA subjects (p=0.005); (2) those who had less years of completed education (p=0.028); (3)

Table I. Variables predicting rheumatoid arthritis (RA) onset and subsequent survival in the generalized structural equation model (GSEM).

Outcome 1:	Variables	OR	OR (95% CI)	*p*-value
incident RA onset	Family history of RA (yes)	7.2	2.3–22.4	0.001
	Rheumatoid factor immunoglobulin M (IgM)	2.0	1.4–2.8	<0.001
	Cigarette smoking daily (ref: non-smoker):			
	1-10 cigarettes	0.9	0.2–4.1	0.855
	11-20 cigarettes	2.5	1.0–6.0	0.042
	21+ cigarettes	5.0	1.8–13.6	0.002
	Matching variable (based on age, sex, and race)		Constrained	
Outcome 2:	**Variables**	**HR**	**HR (95% CI)**	***p*-value**
Mortality (Increased)	Cohort entry age (single year): all subjects	1.1	1.1–1.2	<0.001
	Cigarette smoking (ref: non-smoker):			
	1-10 cigarettes	1.4	0.7–3.1	0.363
	11-20 cigarettes	1.8	1.1–2.9	0.014
	21+ cigarettes	2.1	1.2–3.7	0.013
	Less education completed (years)	1.1	1.0–1.2	0.028
	RA onset (yes) (ref: non-RA)	2.2	1.3–3.8	0.005
	Interaction: RA onset and age (ref: onset age<60 yrs):			
	RA onset and age 60+ years[1]	0.4	0.2–1.0	0.041
	Non-RA onset and age 60+ years	1.0	0.6–1.7	0.957
	Total androgenic-anabolic (A-A) steroids (latent)[2,3]	2.6	1.6–4.1	<0.001
	Interaction: Onset age and A-A steroids (latent):			
	RA onset age<60 years		Constrained	
	RA onset age 60+ years[4]	2.0	1.2-3.3	0.005
	Immunoreactive proteins (latent)	1.3	1.0–1.8	0.033

CI: confidence interval; HR: hazard ratio; OR: odds ratio; RA: rheumatoid arthritis.
[1] RA *vs*. CN relative mortality was less among RA cases with onset 60+ years *vs*. RA onset <60 years.
[2] All biomarkers indicate an increase of one Z-score, except androgenic-anabolic steroids (decrease of one Z-score).
[3] Total mortality was lower with increased androgenic-anabolic steroids z-score (later found in CN onset <60 yrs).
[4] Greater androgenic-anabolic steroids did not reduce mortality in RA subjects with onset <60 *vs*. 60+ years.

those who smoked 11–20 cigarettes/day (p=0.014), and 21+ cigarettes/day (p=0.013) at baseline, and (4) those who had higher levels of baseline entry immunoreactive proteins (p=0.033). In contrast, total mortality was decreased with higher cohort entry androgenic-anabolic steroids (p<0.001) (Table I, Fig. 2). Rheumatoid factor increased incident RA onset (p<0.001, Table I), but did not affect survival (Fig. 2). The Figure 2 GSEM graphic excludes the non-significant matching variable of sex and the initially suspected baseline hormonal variables of luteinising hormone (LH) and prolactin (PRL), which did not correlate with incident RA or survival time.

Notably, in the GSEM Figure 2 graphic, matching variable of age, dichotomised at RA onset <60 *vs*. 60+ years significantly (p=0.041) interacted with survival time. Further analysis (Fig. 3) revealed 65.7% overall mortality in 35 RA cases having onset less than age 60 years compared to 37.1% mortality in 140 matched CN (p=0.002). Mortality was comparable (p=0.997) in 19 cases having RA onset at age 60 years or older (84.2%) *vs*. 76 matched CN (81.6%). In non-GSEM analyses, total dead in

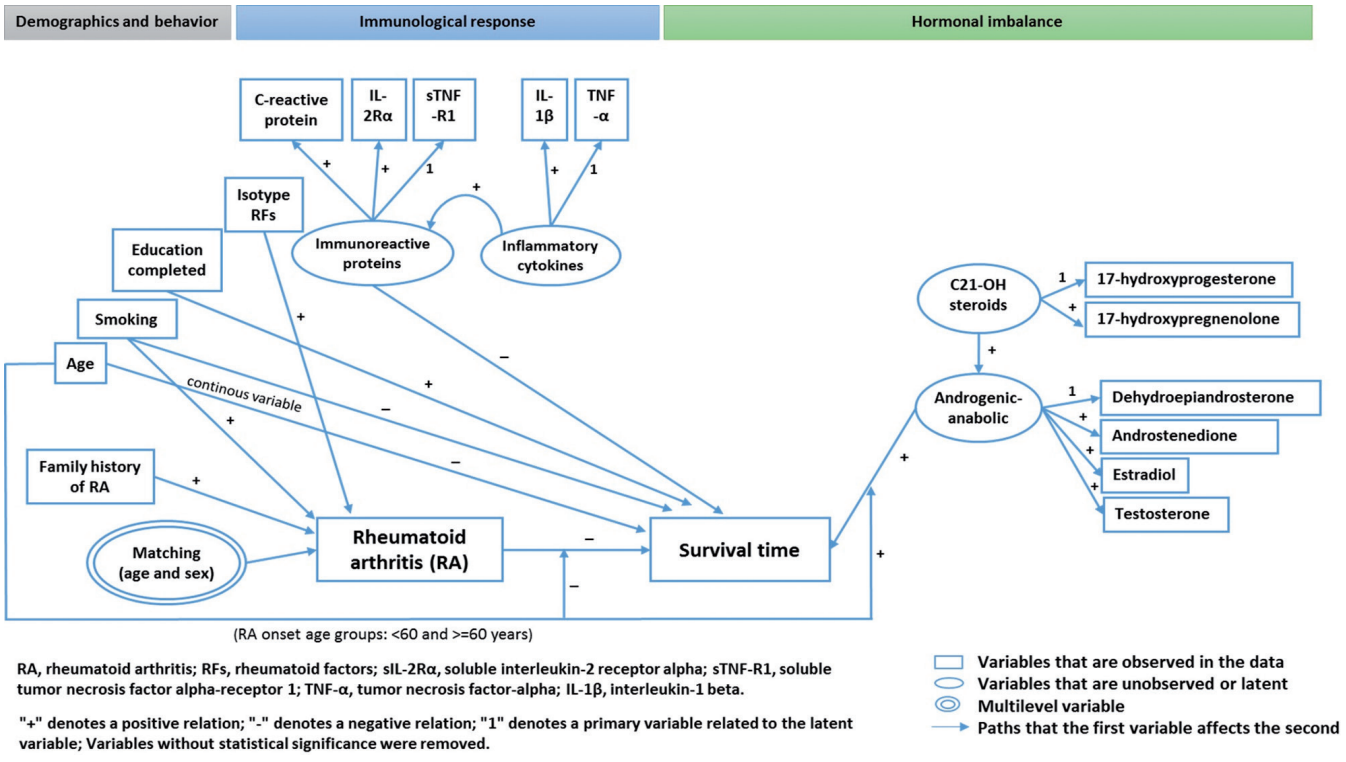

Fig. 2. Significant results of structural equation model for incident rheumatoid arthritis onset and survival.

Table II. Survival in rheumatoid arthritis cases and matched control subjects.

A 30-yr survival in RA cases having onset <60 yrs *vs*. matched CN subjects	Study groups (onset <60 yrs) Cohort mean ages (±SD): Entry 37±8; Onset 48±7	Total (n)	Alive (n, %)	Dead (n, %)	*p*-value (*vs*. CN)
	Total CN (Matched)	140	94 (67.0)	46 (33.0)*	Reference
	RA1 (Good Rx Response)	13	10 (76.9)	3 (23.1)†	0.552
	RA2 (Fair/Limited)	22	3 (13.6)	19 (86.4)†	< 0.001
	Total RA	35	13 (37.1)	22 (62.9)	0.002
B 20-yr survival in RA cases having onset 60+ yrs *vs*. matched CN subjects	Study groups (onset 60+ yrs) Cohort mean ages (± SD): Entry 54±7; Onset 68±6	Total (n)	Alive (n, %)	Dead (n, %)	*p*-value (vs CN)
	Total CN (Matched)	76	30 (39.5)	46 (60.5)*	Reference
	RA1 (Good Rx Response)	6	4 (66.7)	2 (33.3)$	0.226
	RA2 (Fair/Limited)	13	3 (23.1)	10 (76.9)$	0.357
	Total RA	19	7 (36.8)	12 (63.2)	1.000
C Combined 30-yr survival in RA with onset <60 yrs and 20-yr survival with onset 60+ yrs *vs*. total matched CN subjects	Study groups (all onset yrs) Cohort mean ages (± SD): Entry 43±11; Onset 55±12	Total (n)	Alive (n, %)	Dead (n, %)	*p*-value (vs CN)
	Total CN (Matched)	216	124 (57.4)	92 (42.6)‡	Reference
	RA1 (Good Rx Response)	19	14 (73.7)	5 (26.3)#	0.226
	RA2 (Fair/Limited)	35	6 (17.1)	29 (82.9)#	< 0.001
	Total RA	54	20 (37.0)	34 (63.1)‡	0.009

* $p<0.001$, Dead in CN with onset <60 *vs*. 60+ years.
† $p<0.001$, Dead in RA cases with onset <60 yrs, comparing good *vs*. fair/limited Rx responses.
‡ $p=0.009$, Dead in total RA cases *vs*. total CN subjects.
$ $p=0.129$, Dead in RA cases with onset 60+ yrs, comparing good *vs*. fair/limited Rx responses.
# $p<0.001$, Dead in RA cases with all onset ages, comparing good *vs*. fair/limited Rx responses.

the specified observation periods was significantly ($p=0.009$) greater in the 54 RA (63.1%) *vs*. 216 CN (42.6%) (Table II). The mortality difference was significant ($p=0.002$) only in the larger sample of earlier onset RA *vs*. CN subjects (62.9%, n=35 *vs*. 33.0%, n=140, respectively), but not ($p=1.000$) in the smaller sample of later onset subjects (63.2%, n=19 *vs*. 60.5%, n=76, respectively). In 35 RA cases having onsets <60 years, deaths were significantly ($p<0.001$) greater in 22 having fair/limited (82.9%) *vs*. 13 good (23.1%) course-wide therapeutic responses (Table II). The frequency difference in deaths by therapy responses in the small sample (n=19) of later onset RA cases was not significant ($p=0.129$) (Table II). In CN subjects, deaths were highly ($p<0.001$) significantly greater in the 76 later (60.5%) *vs*. 140 earlier (33.0%) onset groups. In contrast, no difference in deaths was found in RA subjects between the 35 having earlier onsets (62.9%) *vs*. 19 having later onsets (63.2%). The major contribution to the increased deaths in total RA *vs*. CN ($p<0.009$) was derived from the earlier onset subjects ($p=0.002$). The difference was particularly significant ($p<0.001$) between the 22 earlier onset RA cases who had fair or limited therapeutic response (86.4%) *vs*. the 140 earlier onset CN subjects (33.0%).

Significant ($p<0.001$) interaction was found between total subjects' increase in androgenic-anabolic *z*-scores and decreased mortality (Table I). Survival from cohort entry age (Y-axis) is positively correlated with entry androgenic-anabolic steroid *z*-scores (X-axis) only in the earlier onset CN, but not in the later CN subjects nor in either RA case subgroup (interaction $p=0.005$, Fig. 4). CLUE age did not significantly interact with androgenic-anabolic steroid *z*-scores in the earlier onset CN subjects ($p=0.392$), indicating that age was not responsible for the relation (or direction) between increasing androgenic-anabolic steroid *z*-scores and survival in the earlier onset CN subjects (Fig. 4). The significant relation between increasing androgenic-anabolic steroid *z*-scores and survival was specific to the earlier onset CN subjects (Fig. 4).

Discussion

Although multidimensional or structural equation model has been applied to analysing function (32), fatigue (33) and quality of life (34) in RA patients, this study is the first to our knowledge in which GSEM is used to analyse risk factors for both RA onset and its subsequent survival. Generalized SEM utilises regression analysis to express association among variables and to correlate groups of interrelated variables into a single factor or latent construct involving causal path analysis which can point towards a dependent outcome variable (14). When considering causal pathways, this model also permits graphic representation where direct effects are depicted as vectors stemming from independent risk factors or exposure variables pointing towards a dependent outcome variable. In this study, GSEM analysed both out-

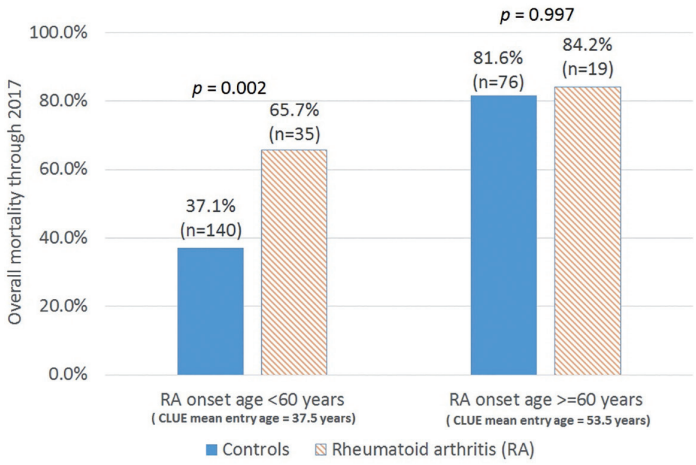

Fig. 3. Mortality of rheumatoid arthritis *vs*. control subjects by RA onset ages <60 *vs*. 60+ years.

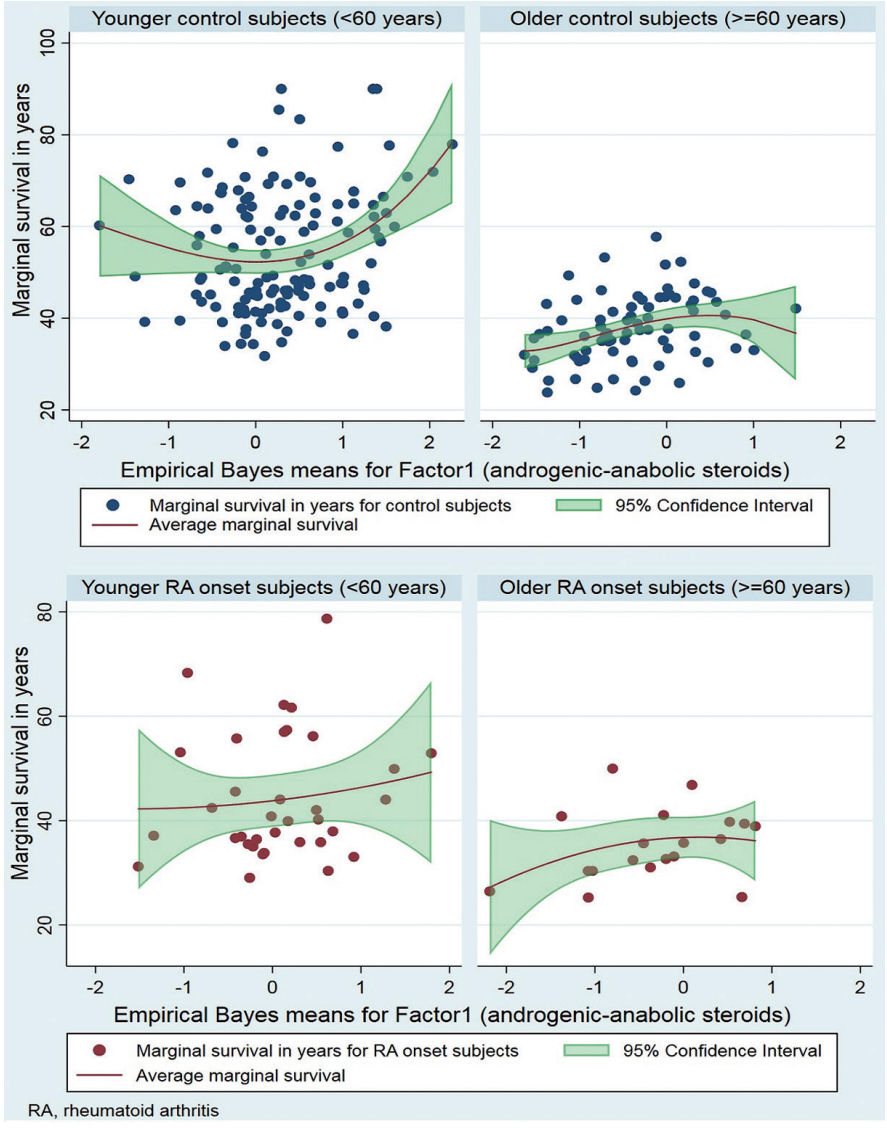

Fig. 4. Relation between androgenic-anabolic steroids *z*-scores and survival of rheumatoid arthritis and control subjects having onset ages <60 *vs*. 60+ years.

come variables of incident RA and its subsequent survival (Figs. 1-2). Such dichotomous outcome analysis is more complex and challenging when including survival-time, categorical factors, and/or censored data (Figs. 1-2). In our study, a GSEM was used to cope with these challenges which have not been addressed in previous studies of RA patients (32-34).

Generalized SEM is often used as both a confirmatory tool of previously established empirical associations and an exploratory method (14-17). From our data set, we first outlined a hypothesised theoretical model of factors influencing onset risk and survival of RA (Fig. 1), from which significant GSEM interrelations were derived (Fig. 2). Increased risk of incident RA was associated with a positive family history, baseline cigarette smoking and IgM RF (Table I), which were characterised in previous reports from this cohort (9, 11, 24, 35) and in the literature (36, 37). Baseline factors which increased total mortality or decreased survival time included RA, subject entry age, cigarette smoking, and immunoreactive proteins. Factors which increased total survival time included years of completed education and androgenic-anabolic steroid levels (Table I), operating only in the younger onset CN subjects (Fig. 4). Impacts of RA, age, smoking, and immunoreactive proteins on diminished longevity have been reported (37-40), but this GSEM study supports these causal pathways in a single model (Fig. 2).

Dichotomised RA onset age, assigned to matched CN within 54 sets, was found to mediate several relations. Earlier onset (<60 yrs.) significantly decreased survival time of RA vs CN, which was not found in later onset (60+ yrs.) subjects (Fig. 3). Recently, a cohort study in Canada found potential life years lost (before the age of 75) among RA patients was about twice that of persons without RA (41). A retrospective study of 130 hospital-ascertained RA patients from Leiden, The Netherlands, determined clinical and mortality outcomes by RF status and differential disease onset ages at initial diagnosis, conventionally defined as younger (<60 yrs.) and older (60+ yrs.) (42). All 61 seronegative patients diagnosed between 1980–86 were included in study and a randomised sample of 69 seropositive patients. At a mean of 5.6 years after diagnoses, standardised mortality ratios (95% CI) were 2.79 (1.70–4.14) for to-

tal 69 seropositive and 0.45 (0.08–1.13) for total 61 seronegative patients (42). The authors concluded presence of RF at diagnosis determines outcome more than onset age, although numbers of deaths (expected) were few in the 56 younger onset patients, 3 (0.84) in 28 RF positive and 0 (1) in 28 RF negative patients for comparison with the older onset RA cases. In the current study, baseline IgM RF predicted RA onset, but was not associated with decreased survival, unlike immunoreactive proteins (Table I, Fig. 2).

A US sample of 779 RA patients were followed (mean 2.5, range 0.1-6.3 years) to assess a global disease severity scale (<2, 2-4, 5+) as a predictor of 75 (9.6%) deaths (43). Estimated deaths (percentage dead) increased by progressive disease severity scale: <2 (n=257), 9 deaths (4%); 2–4 (n=270), 28 deaths (10%), and 5+ (n=252), 38 deaths (15%) (Fig. 3B in 43). The latter mortality trend (43) is consistent with a preceding survival analysis of 3 therapy response categories in this CLUE cohort (21) and mortality in this study (Table II). Further research is needed to determine mechanisms of how specific and reliable components of therapy response or disease severity may independently influence mortality in RA.

The GSEM also showed a significant positive correlation between increased baseline androgenic-anabolic steroid Z-scores and greater survival in total subjects (Fig. 2), which derived only from CN having onset <60 years (Fig. 4). The latter cohort subgroup had mean (±SD) baseline entry age of 37(±8) and onset of 48(±7). In a population-based cohort (1972–74) of 242 older white men aged 50 to 79 years residing in Rancho Bernardo, California, an inverse significant (p<0.050) association was found between the single baseline dehydroepiandrosterone sulfate (DHEAS) level and subsequent12-year mortality for all causes and cardiovascular disease (44). In a subsequent southern France cohort (entry 1988) of subjects 65+ years of age, 8-year mortality was inversely related to baseline DHEAS level in men under 70 years (p=0.007), but not significant in women (45). No reports were found of relation of baseline DHEAS levels to subsequent mortality in cohorts less than 50 years and no relation was found between serum androstenedione, testosterone, or estradiol levels and subsequent mortality. The specific relation of the androgenic-anabolic latent variable in this study to serum DHEAS is not known, although DHEA was hypothesised to be the primary steroid among those contributing to the latent variable. The relation of androgenic-anabolic steroid and DHEAS levels in persons less than 50 years to subsequent mortality from all causes and cardiovascular disease deserves further study.

Limitations of our study need to be noted. First, the sample size is modest for a complex SEM. Second, biomarkers in our analysed data were only measured at baseline, excluding our model from examining impact of change of biomarkers over time on risk of incident RA and its survival. Third, the potential of unidentified confounding effect exists in cohort studies, although reduced by close age and sex matching in this community-based population (46). An asset of this study is its unique design of a 43-year cohort which lends feasibility to analysing predictors of both incident RA and long-term survival. This study used GSEM in a confirmatory and exploratory manner. The confirmatory results in this study support GSEM's feasibility in analyses of longitudinal data assessing the new onset and long-term survival of RA. Marginal structural models have also been introduced for causal inference in epidemiology in order to appropriately control for the effects of time-varying confounders that are affected by prior treatment (exposure) (47, 48). A further marginal structural model could be considered to investigate the effect of RA and its responses to therapy upon survival controlling for effects of confounders, including time-varying factors (*e.g.* age).

Conclusions

This is the first report to our knowledge of GSEM applied to an established RA cohort to investigate risk factors for incident RA onset and its subsequent survival. The multiple confirmatory and novel results of this analytic method support its application as a promising approach towards elucidating complex relations and causal pathways influencing RA onset and its long-term mortality. The GSEM identified the novel interaction between ages at earlier and later onsets of RA and survival, which led to documentation of a mortality hazard from insufficient therapy responses, particularly in the younger onset subjects. The significant increase of baseline androgenic-anabolic steroids z-scores with greater survival only in earlier onset CN subjects is novel and also deserves further study.

Acknowledgement

We would like to acknowledge medical student David X. Gao for his technical contributions in the literature review.

References

1. GIBOFSKY A: Epidemiology, pathophysiology, and diagnosis of rheumatoid arthritis: A Synopsis. *Am J Manag Care* 2014; 20 (7 Suppl.): S128-35.
2. HUNTER TM, BOYTSOV NN, ZHANG X, SCHROEDER K, MICHAUD K, ARAUJO AB: Prevalence of rheumatoid arthritis in the United States adult population in healthcare claims databases, 2004-2014. *Rheumatol Int* 2017; 37: 1551-57.
3. SCOTT DL, WOLFE F, HUIZINGA TW: Rheumatoid arthritis. *Lancet* 2010; 376: 1094-108.
4. RASCH EK, HIRSCH R, PAULOSE-RAM R, HOCHBERG MC: Prevalence of rheumatoid arthritis in persons 60 years of age and older in the United States: effect of different methods of case classification. *Arthritis Rheum* 2003; 48: 917-26.
5. BIRNBAUM H, PIKE C, KAUFMAN R, MARYNCHENKO M, KIDOLEZI Y, CIFALDI M: Societal cost of rheumatoid arthritis patients in the US. *Curr Med Res Opin* 2010; 26: 77-90.
6. GIBOFSKY A: Overview of epidemiology, pathophysiology, and diagnosis of rheumatoid arthritis. *Am J Manag Care* 2012; 18 (13 Suppl.): S295-302.
7. GORONZY JJ, SHAO L, WEYAND CM: Immune aging and rheumatoid arthritis. *Rheum Dis Clin North Am* 2010; 36: 297-310.
8. VERMA MK, SOBHA K: Understanding the major risk factors in the beginning and the progression of rheumatoid arthritis: current scenario and future prospects. *Inflamm Res* 2015; 64: 647-59.
9. MASI AT, REHMAN AA, ELMORE KB, ALDAG JC: Serum acute phase protein and inflammatory cytokine network correlations: comparison of a pre-rheumatoid arthritis and non-rheumatoid arthritis community cohort. *J Innate Immun* 2013; 5: 100-13.
10. MASI AT, REHMAN AA, CUTOLO M, ALDAG JC: Do women with premenopausal-onset rheumatoid arthritis have relative insufficiency or imbalance of adrenocortical steroids? *Ann N Y Acad Sci* 2014; 1317: 7-16.

11. MASI AT, REHMAN AA, JORGENSON LC, ALDAG JC: Preclinical biomarker associations with both incident rheumatoid arthritis and its subsequent mortality: sex effects in a 41-year, community-based, case-control cohort study. *Clin Exp Rheumatol* 2017; 35: 966-74.
12. JOSEPH RM, MOVAHEDI M, DIXON WG, SYMMONS DP: Smoking-related mortality in patients with early rheumatoid arthritis: a retrospective cohort study using the clinical practice research datalink. *Arthritis Care Res* 2016; 68: 1598-606.
13. DADONIENE J, STROPUVIENE S, STUKAS R, VENALIS A, SOKKA-ISLER T: Predictors of mortality in patients with rheumatoid arthritis in Lithuania: Data from a cohort study over 10 years. *Medicina* (Kaunas) 2015; 51: 25-31.
14. ROMAN-URRESTARAZU A, ALI FM, REKA H, RENWICK MJ, ROMAN GD, MOSSIALOS E: Structural equation model for estimating risk factors in type 2 diabetes mellitus in a Middle Eastern setting: evidence from the STEPS Qatar. *BMJ Open Diabetes Res Care* 2016; 4: e000231.
15. JALALI-FARAHANI S, AMIRI P, KARIMI M, GHARIBZADEH S, MIRMIRAN P, AZIZI F: Socio-behavioral factors associated with overweight and central obesity in tehranian adults: a structural equation model. *Int J Behav Med* 2017; 24: 110-19.
16. VERKUILEN J: Explanatory Item Response Models: A Generalized Linear and Nonlinear Approach by P. de Boeck and M. Wilson and Generalized Latent Variable Modeling: Multilevel, Longitudinal and Structural Equation Models by A. Skrondal and S. Rabe-Hesketh. *Psychometrika* 2006; 71: 415-8.
17. KUPEK E: Beyond logistic regression: structural equations modelling for binary variables and its application to investigating unobserved confounders. *BMC Med Res Methodol* 2006; 6: 13.
18. STATA: Generalized SEM. [cited 2017 August 31]; Available from: https://www.stata.com/features/overview/generalized-sem/.
19. GALLICCHIO L, CHANG HH, CHRISTO DK *et al*.: Single nucleotide polymorphisms in obesity-related genes and all-cause and cause-specific mortality: a prospective cohort study. *BMC Med Genet* 2009; 10: 103.
20. GALLICCHIO L, KALESAN B, HUANG HY, STRICKLAND P, HOFFMAN SC, HELZLSOUER KJ: Genetic polymorphisms of peroxisome proliferator-activated receptors and the risk of cardiovascular morbidity and mortality in a community-based cohort in Washington County, Maryland. *PPAR Res* 2008; 2008: 276581.
21. MASI AT, REHMAN AA, JORGENSON LC, ALDAG JC: Increased mortality of incident rheumatoid arthritis versus matched non-RA control subjects: a community-based long-term prospective cohort study. *Clin Exp Rheumatol* 2017; 35: 277-87.
22. MASI AT, REHMAN AA, JORGENSON LC, SMITH JM, ALDAG JC: Sexual dimorphisms of adrenal steroids, sex hormones, and immunological biomarkers and possible risk factors for developing rheumatoid arthritis. *Int J Endocrinol* 2015; 2015: 929246.
23. MASI AT, REHMAN AA, CHATTERTON RT *et al*.: Controlled cohort study of serum gonadal and adrenocortical steroid levels in males prior to onset of rheumatoid arthritis (pre-RA): a comparison to pre-RA females and sex differences among the study groups. *Int J Rheumatol* 2013; 2013: 284145.
24. MASI AT, ELMORE KB, REHMAN AA, CHATTERTON RT, GOERTZEN NJ, ALDAG JC: Lower serum androstenedione levels in pre-rheumatoid arthritis versus normal control women: correlations with lower serum cortisol levels. *Autoimmune Dis* 2013; 2013: 593493.
25. GERLAG DM, RAZA K, VAN BAARSEN LG *et al*.: EULAR recommendations for terminology and research in individuals at risk of rheumatoid arthritis: report from the Study Group for Risk Factors for Rheumatoid Arthritis. *Ann Rheum Dis* 2012; 71: 638-41.
26. ARNETT FC, EDWORTHY SM, BLOCH DA *et al*.: The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. *Arthritis Rheum* 1988; 31: 315-24.
27. WOLF EJ, HARRINGTON KM, CLARK SL, MILLER MW: Sample size requirements for structural equation models: an evaluation of power, bias, and solution propriety. *Edu Psychol Meas* 2013; 76: 913-34.
28. SIDDIQUI K: Heuristics for sample size determination in multivariate statistical techniques. *Word Applied Sciences Journal* 2013; 27: 3.
29. HOYLE RH, GOTTFREDSON NC: Sample size considerations in prevention research applications of multilevel modeling and structural equation modeling. *Prev Sci* 2015; 16: 987-96.
30. YAMAGUCHI Y, MISUMI T, MARUO K: A comparison of multiple imputation methods for incomplete longitudinal binary data. *J Biopharm Stat* 2018; 28: 645-67.
31. BERAN TN, VIOLATO C: Structural equation modeling in medical research: a primer. *BMC Res Notes* 2010; 3: 267.
32. ESCALANTE A, HAAS RW, DEL RINCON I: A model of impairment and functional limitation in rheumatoid arthritis. *BMC Musculoskeletal Disord* 2005; 6: 16.
33. NICASSIO PM, ORMSETH SR, CUSTODIO MK, IRWIN MR, OLMSTEAD R, WEISMAN MH: A multidimensional model of fatigue in patients with rheumatoid arthritis. *J Rheumatol* 2012; 39: 1807-13.
34. SANTOS EJF, DUARTE C, FERREIRA RJO, PINTO AM, GEENEN R, DA SILVA JAP: Determinants of happiness and quality of life in patients with rheumatoid arthritis: a structural equation modelling approach. *Ann Rheum Dis* 2018; 77: 1118-24.
35. MASI AT, ALDAG JC: Integrated neuroendocrine immune risk factors in relation to rheumatoid arthritis: should rheumatologists now adopt a model of a multiyear, presymptomatic phase? *Scand J Rheumatol* 2005; 34: 342-52.
36. BAKA Z, BUZAS E, NAGY G: Rheumatoid arthritis and smoking: putting the pieces together. *Arthritis Res Ther* 2009; 11: 238.
37. DI GIUSEPPE D, DISCACCIATI A, ORSINI N, WOLK A: Cigarette smoking and risk of rheumatoid arthritis: a dose-response meta-analysis. *Arthritis Res Ther* 2014; 16: R61.
38. DEON D, AHMED S, TAI K *et al*.: Cross-talk between IL-1 and IL-6 signaling pathways in rheumatoid arthritis synovial fibroblasts. *J Immunol* 2001; 167: 5395-403.
39. GABRIEL SE: Why do people with rheumatoid arthritis still die prematurely? *Ann Rheum Dis* 2008; 67 (Suppl. 3): iii30-34.
40. PIRIE K, PETO R, REEVES GK, GREEN J, BERAL V: The 21st century hazards of smoking and benefits of stopping: a prospective study of one million women in the UK. *Lancet* 2013; 381: 133-41.
41. WIDDIFIELD J, PATERSON JM, HUANG A, BERNATSKY S: Causes of death in rheumatoid arthritis: How do they compare to the general population? *Arthritis Care Res* 2018; 70: 1748-55.
42. VAN SCHAARDENBURG D, HAZES JM, DE BOER A, ZWINDERMAN AH, MEIJERS KA, BREEDVELD FC: Outcome of rheumatoid arthritis in relation to age and rheumatoid factor at diagnosis. *J Rheumatol* 1993; 20: 45-52.
43. NAVARRO-CANO G, DEL RINCON I, POGOSIAN S, ROLDAN JF, ESCALANTE A: Association of mortality with disease severity in rheumatoid arthritis, independent of comorbidity. *Arthritis Rheum* 2003; 48: 2425-33.
44. BARRETT-CONNOR E, KHAW KT, YEN SS: A prospective study of dehydroepiandrosterone sulfate, mortality, and cardiovascular disease. *N Engl J Med* 1986; 24: 1519-24
45. MAZAT L, LAFONT S, BERR C *et al*.: Prospective measurements of dehydroepiandrosterone sulfate in a cohort of elderly subjects: relationship to gender, subjective health, smoking habits, and 10-year mortality. *Proc Natl Acad Sci USA* 2001; 98: 8145-50.
46. MAMDANI M, SYKORA K, LI P *et al*.: Reader's guide to critical appraisal of cohort studies: 2. Assessing potential for confounding. *BMJ* 2005; 330: 960-62.
47. ROBINS JM, HERNAN MA, BRUMBACK B: Marginal structural models and causal inference in epidemiology. *Epidemiology* 2000; 11: 550-60.
48. MOERKERKE B, LOEYS T, VANSTEELANDT S: Structural equation modeling versus marginal structural modeling for assessing mediation in the presence of posttreatment confounding. *Psychol Methods* 2015; 20: 204-20.