

# Smoking Is Associated with Higher Disease Activity in Rheumatoid Arthritis: A Longitudinal Study Controlling for Time-varying Covariates

Milena A. Gianfrancesco , Laura Trupin , Stephen Shiboski , Mark van der Laan, Jonathan Graf, John Imboden, Jinoos Yazdany , and Gabriela Schmajuk 

**ABSTRACT. Objective.** Prior studies around the relationship between smoking and rheumatoid arthritis (RA) disease activity have reported inconsistent findings, which may be ascribed to heterogeneous study designs or biases in statistical analyses. We examined the association between smoking and RA outcomes using statistical methods that account for time-varying confounding and loss to followup.

**Methods.** We included 282 individuals with an RA diagnosis using electronic health record data collected at a public hospital between 2013 and 2017. Current smoking status and disease activity were assessed at each visit; covariates included sex, race/ethnicity, age, obesity, and medication use. We used longitudinal targeted maximum likelihood estimation to estimate the causal effect of smoking on disease activity measures at 27 months, and compared results to conventional longitudinal methods.

**Results.** Smoking was associated with an increase of 0.64 units in the patient global score compared to nonsmoking ( $p = 0.01$ ), and with 2.58 more swollen joints ( $p < 0.001$ ). While smoking was associated with a higher clinical disease activity score (2.11), the difference was not statistically significant ( $p = 0.22$ ). We found no association between smoking and physician global score, or C-reactive protein levels, and an inverse association between smoking and tender joint count ( $p = 0.05$ ). Analyses using conventional methods showed a null relationship for all outcomes.

**Conclusion.** Smoking is associated with higher levels of disease activity in RA. Causal methods may be useful for investigations of additional exposures on longitudinal outcome measures in rheumatologic disease. (J Rheumatol First Release December 1 2018; doi:10.3899/jrheum.180262)

## Key Indexing Terms:

RHEUMATOID ARTHRITIS      SMOKING      DISEASE ACTIVITY      EPIDEMIOLOGY

From the Division of Rheumatology, Department of Medicine, University of California, San Francisco, California, USA.

This work was supported by the US National Institute of Arthritis and Musculoskeletal and Skin Diseases of the National Institutes of Health (NIH; grant number F32 AR070585 to MG and K23 AR063770 to GS) and the Agency for Healthcare Research and Quality (grant number R01 HS024412 to JY). Drs. Graf, Yazdany, and Schmajuk are also supported by the Russell/Engleman Medical Research Center for Arthritis. The content is solely the responsibility of the authors and does not necessarily represent the official views of the Agency for Healthcare Research and Quality or the NIH.

M.A. Gianfrancesco, PhD, MPH, Division of Rheumatology, Department of Medicine, University of California, San Francisco; L. Trupin, MPH, Division of Rheumatology, Department of Medicine, University of California, San Francisco; S. Shiboski, PhD, Department of Epidemiology and Biostatistics, University of California, San Francisco; M. van der Laan, PhD, Division of Biostatistics, School of Public Health, University of California, Berkeley; J. Graf, MD, Division of Rheumatology, Department of Medicine, University of California, San Francisco; J. Imboden, MD, Division of Rheumatology, Department of Medicine, University of California, San Francisco; J. Yazdany, MD, MPH, Division of Rheumatology, Department of Medicine, University of California, San Francisco; G. Schmajuk, MD, MS, Division of Rheumatology, Department of Medicine, University of California, San Francisco, and Veterans Affairs Medical Center, San Francisco, California, USA.

Address correspondence to Dr. M.A. Gianfrancesco, Division of Rheumatology, Department of Medicine, University of California, San Francisco; 513 Parnassus Ave., San Francisco, California 94143, USA. E-mail: Milena.Gianfrancesco@ucsf.edu

Accepted for publication August 10, 2018.

Rheumatoid arthritis (RA) is a chronic inflammatory disease that commonly leads to functional limitation and physical disability<sup>1</sup>. Smoking is considered one of the strongest environmental risk factors for RA. A metaanalysis demonstrated that the odds of developing RA for smokers were nearly 2-fold compared to nonsmokers<sup>2</sup>.

The relationship between smoking and RA outcomes is less clear. While previous studies have shown strong associations between smoking and nodule formation<sup>3,4,5,6,7,8</sup>, others have shown conflicting results. For example, Matthey, *et al* showed a significant increase in radiographic damage in smokers compared to nonsmokers<sup>5</sup>, while other studies found no association<sup>9,10</sup>, including a longitudinal study of over 2000 individuals<sup>10</sup>.

Inconsistent findings in disease activity measures may be ascribed to heterogeneous study designs, measurement error in key variables, or biases in statistical analyses. Observational studies may be biased in sampling, and cross-sectional studies may not be able to assess the timing or “dose” of the exposure and the outcome. Within longitudinal study designs, improper adjustment for time-varying variables (i.e., changes in exposure status or covariates over time) can lead to biased results, specifically when a covariate serves as both a

confounder and an intermediate variable. Standard regression techniques may lead to biased estimates in this context<sup>11</sup>, but other methods, such as inverse probability weighted (IPW) estimators<sup>12</sup> and longitudinal targeted maximum likelihood estimation (LTMLE)<sup>13</sup>, can yield unbiased estimates that also have a valid causal interpretation given certain assumptions.

We used LTMLE, a doubly robust, maximum-likelihood-based estimator that accounts for time-varying covariates and potentially informative missingness, to examine the association between current smoking status and RA disease activity at 27 months. We also conducted analyses using conventional methods as a comparison.

## MATERIALS AND METHODS

**Study population.** We included individuals with at least 2 rheumatology clinic visits within 12 months, each coded with a diagnosis of RA (International Classification of Diseases, 9th ed diagnosis code: 714.0) between January 1, 2013, and July 30, 2017, from the electronic health records (EHR) of a public hospital (n = 391). We restricted the sample to patients who had at least 2 observations of one of the disease activity measures analyzed (n = 307); 43 individuals did not have any outcomes recorded, while 41 had only 1 disease activity measure documented. We also excluded 16 individuals with “uninformative” datapoints (i.e., only 2 visits, within 30 days, because these patients had inadequate longitudinal followup) and 9 individuals with missing body mass index values that could not be recovered in chart review (final n = 282).

Observation windows during the study period were defined in 3-month intervals, which is the standard followup interval for patients with RA in our health system. If there were no available data for the 3-month interval, we expanded the observation window to  $\pm 30$  days; otherwise, the observation was designated as a missed visit. The exposure (current cigarette smoking, yes/no) was assessed at each timepoint through EHR for each clinic visit. Outcomes included disease activity measured by the Clinical Disease Activity Index (CDAI)<sup>14</sup> and its components: patient’s global assessment (PtGA), physician’s global assessment (PGA), and 28 tender (TJC) and swollen joint counts (SJC). We also examined C-reactive protein (CRP) laboratory values. Laboratory values that included a minimum or maximum value (e.g., “< 3” or “> 200”) were replaced with the number (e.g., “3,” “200”). In our sample, laboratory values were generally assessed on the same day of a patient’s clinic visit, and therefore can be directly linked with the EHR visit data. If multiple values were recorded for a disease activity measure within a 3-month interval, an average of the scores was used. In circumstances where outcomes were missing, we expanded the observation window to  $\pm 30$  days; if the outcome was still missing, we imputed using last value carried forward and generated a variable to indicate whether the outcome value was imputed to record any informative missingness.

Baseline covariates included sex, race/ethnicity (white non-Hispanic, black non-Hispanic, Asian non-Hispanic, Hispanic), age at first observation, smoking status prior to index date (ever/never), obesity status (body mass index  $> 30$  kg/m<sup>2</sup>), rheumatoid factor (RF) status, and anticyclic citrullinated peptide (anti-CCP) status. Disease-modifying antirheumatic drugs (DMARD) were assessed at each timepoint and coded as synthetic DMARD (yes/no) or biologic/small molecule DMARD (yes/no; Supplementary Table 1, available with the online version of this article). Current smoking (exposure of interest) and disease activity measures (outcome) were measured at baseline and each timepoint.

Patients were observed from baseline (the second of 2 encounters within 12 months) until loss to followup, nine 3-month intervals (following the baseline observation), or on July 31, 2017, whichever occurred first. We chose a maximum study period of 27 months of observation because the proportion of patients followed longer than 27 months dropped sharply ( $> 40\%$  censored). Continuing the study period would have reduced statistical power.

**Statistical analyses.** Within longitudinal study designs, estimation of exposure effects based on improper adjustment for time-varying confounding variables can lead to biased results, specifically when a covariate serves as both a confounder and an intermediate variable for an exposure of interest<sup>11</sup>. An example of this is medication use, which serves as both a confounder and an intermediate variable in the relationship between smoking and disease activity: adjusting for this variable may attenuate the estimated association between exposure and outcome because it lies in the causal pathway as an intermediate variable; but not adjusting for it can lead to biased results because it can still represent a confounder. Conventional regression fails to appropriately account for this type of bias, but other methods, such as IPW<sup>12,15,16</sup> and LTMLE<sup>13</sup>, allow for the estimation with a valid causal interpretation. Similar to the conventional approach, validity relies on assumptions that important confounding influences are identified, measured, and properly accounted for in analyses. In particular, IPW and LTMLE approaches require that the sequential randomization (i.e., conditional on the observed covariate and exposure history, the exposure is assigned completely at random) and positivity (i.e., smokers and nonsmokers are present in each covariate stratum) assumptions are met<sup>17</sup>.

We used LTMLE to estimate the marginal association of current smoking status on disease activity as measured by CDAI, PtGA, PGA, SJC, TJC, and CRP values at 27 months<sup>13</sup>. Our causal quantity of interest is defined using a marginal structural working model to summarize how the mean counterfactual outcome at 27 months varies as a function of the intervention rule (always smoke and not censored vs never smoke and not censored), baseline covariates, and time-varying covariates (Supplementary Data 1, available with the online version of this article). LTMLE estimates both outcome and exposure mechanisms using semiparametric techniques, making the resulting estimates doubly robust against misspecification of either of these mechanisms alone.

In our study, LTMLE estimated the cumulative effect of smoking while accounting for censoring and time-varying covariates, including missed visit status and medication<sup>18</sup>. As expected in EHR data, not all patients were seen every 3 months ( $\pm 30$  days). Rather than censoring at the time of first missed visit, which would substantially reduce the number of individuals in our dataset, LTMLE was used to control for the possibly informative measurement process; that is, to account for missing timepoints that may have been influenced by time-varying covariates. Individuals were considered censored only if they were no longer observed in the clinic at any time during the followup period.

Time-dependent covariates, such as medication and missed clinic visits, may affect both the outcome, as well as censored status. These covariates may also be influenced by the exposure. Therefore, time-dependent confounding is present and standard regression methods are likely to yield biased estimates. Alternate methods, such as LTMLE, allow us to estimate the statistical parameter best approximating our causal parameter of interest. LTMLE analyses were conducted with the “ltmle” package in R<sup>19</sup>.

Chart reviews were conducted for missing data not automatically extracted through EHR tables. For remaining missing values, information was imputed using last value carried forward. Additionally, indicator variables were created to represent whether an individual’s outcome was imputed (e.g., in situations where the patient was present in the clinic, but an outcome value not recorded), and whether an individual missed a visit, but was not yet censored (e.g., patient returned before end of followup). These 2 variables were included as covariates in analyses.

A conventional longitudinal analysis approach for estimation of marginal effects, specifically, generalized estimating equations (GEE)<sup>20</sup>, was also used to model the longitudinal association between current smoking status and disease activity measures over the study period. GEE estimates the expected difference in disease activity for a unit change in the predictor (i.e., comparing smokers to nonsmokers) over the whole population. We modeled the association using an exchangeable working correlation structure and robust standard errors. Non-normally distributed outcomes, such as CDAI, PGA, and CRP values, were also modeled with square root-transformed values.  $\beta$  coefficients were estimated with smoking as the primary predictor of each multivariate model adjusted for appropriate covariates.

Our exposure of interest, current smoking status, and outcomes (disease activity as measured by CDAI, PtGA, PGA, SJC, TJC, and CRP level) were measured at baseline and each timepoint. Analyses using LTMLE and GEE controlled for baseline covariates including sex, race/ethnicity, age, obesity (yes/no), previous smoking status (ever/never), RF status, anti-CCP status, and baseline disease activity. Time-varying covariates included synthetic DMARD, biologic/small molecule DMARD, whether the outcome value was imputed, and missed visit status. LTMLE results additionally accounted for censored status. Analyses were conducted in R (v.3.3.1) or Stata (v.15.0). We created 95% CI and conducted 2-sided hypothesis tests controlling the type I error rate at 5% ( $\alpha = 0.05$ ). The study was approved by the Committee on Human Research at the University of California, San Francisco, California, USA (study number:10-04740).

## RESULTS

*Patient demographic and disease characteristics at baseline.* Demographic and disease characteristics of patients with RA at baseline are described in Table 1. Patients were predominately female (83%), with a mean age of  $59.4 \pm 11.9$  years, and 91% were racial/ethnic minorities. Over 30% of patients were prescribed a biologic or small-molecule DMARD, and 91% were prescribed a synthetic DMARD. About 10% of patients were current smokers at baseline, while 11% reported smoking at some time in the past. The percentage of current smokers over the study period varied between 8 and 11% (data not shown).

There were no significant differences in disease activity measures (CDAI, PtGA, PGA, SJC, TJC, and CRP level) between current smokers compared to nonsmokers at baseline ( $p > 0.05$ , data not shown).

Table 1. Baseline characteristics of patients with rheumatoid arthritis included in the study (n = 282).

Characteristics	N (%) or Mean $\pm$ SD
Female	234 (83)
Age, yrs	59.40 $\pm$ 11.93
Race/ethnicity	
White, non-Hispanic	25 (9)
Asian, non-Hispanic	83 (29)
Black, non-Hispanic	22 (8)
Hispanic	152 (54)
Previous smoker	31 (11)
Current smoker	28 (10)
Obese (BMI $\geq$ 30 kg/m <sup>2</sup> )	106 (38)
Prescribed biologic or small-molecule DMARD	95 (34)
Prescribed synthetic DMARD	256 (91)
RF-positive, n = 280	221 (79)
Anti-CCP, n = 280	249 (89)
Clinical Disease Activity Index (range 0–76)	15.13 $\pm$ 11.61
PtGA (range 0–10)	4.73 $\pm$ 2.75
PGA (range 0–10)	2.71 $\pm$ 2.30
Swollen joint count (0–28)	4.50 $\pm$ 5.29
Tender joint count (0–28)	3.70 $\pm$ 5.45
CRP, mg/l	10.03 $\pm$ 17.63

BMI: body mass index; DMARD: disease-modifying antirheumatic drugs; RF: rheumatoid factor; anti-CCP: anticyclic citrullinated peptide; PtGA: patient's global assessment; PGA: physician's global assessment; CRP: C-reactive protein.

*LTMLE results.* LTMLE results demonstrating the expected mean differences in disease activity scores between current smokers and nonsmokers at 27 months are shown in Table 2. We found a significant marginal effect between current smoking status and PtGA score ( $p = 0.01$ ); current smokers on average had a PtGA score 0.64 units higher than nonsmokers (current smokers = 4.82 vs nonsmokers = 4.19). Additionally, current smokers had on average 2.58 more swollen joints than nonsmokers (current smokers = 5.91 vs nonsmokers = 3.33;  $p < 0.001$ ). Smoking was associated with a 2.11 increase in CDAI score compared to nonsmoking, but the difference was not statistically significant ( $p = 0.22$ ). There was no significant association between current smoking and PGA score ( $p = 0.63$ ) or CRP levels ( $p = 0.61$ ). In contrast, we detected an inverse relationship between current smoking and TJC ( $-1.09$ , 95% CI  $-2.16$  to  $-0.02$ ;  $p = 0.05$ ).

*Conventional longitudinal analysis results.* Table 3 demonstrates the expected difference in disease activity measures over time comparing current smokers to nonsmokers using a conventional longitudinal method, GEE. There were no significant differences found using GEE ( $p > 0.10$ ). Analyses using transformed values for non-normal outcome variables (i.e., CDAI, PGA score, and CRP) demonstrated similar results (data not shown).

## DISCUSSION

We found current smoking status to be associated with higher levels of disease activity as measured by PtGA score and SJC using LTMLE. To our knowledge, this is the first study to examine the longitudinal effect of smoking on disease activity using semiparametric methods that account for time-varying confounding and potentially informative missingness.

Table 2. Longitudinal targeted maximum likelihood estimation results demonstrating the expected mean difference in disease activity scores between smokers and nonsmokers at 27 months.

Variables	Smokers Estimate	Nonsmokers Estimate	Mean Difference (95% CI), p
CDAI	14.77	12.66	2.11 ( $-1.26$ to $5.47$ ), $p = 0.22$
PtGA	4.82	4.19	0.64 ( $0.16$ – $1.12$ ), $p = 0.01$
PGA	2.43	2.23	0.20 ( $-0.61$ to $1.01$ ), $p = 0.63$
Swollen joints	5.91	3.33	2.58 ( $1.28$ – $3.89$ ), $p < 0.001$
Tender joints	1.53	2.62	$-1.09$ ( $-2.16$ to $-0.02$ ), $p = 0.05$
CRP, mg/l	13.22	12.26	0.96 ( $-2.74$ to $4.66$ ), $p = 0.61$

The estimated causal effect of smoking on disease activity at 27 months was adjusted for baseline covariates (sex, age, race/ethnicity, past smoking, disease activity, obesity, RF status, anti-CCP status) and time-varying covariates (biologic/small molecule DMARD, synthetic DMARD, whether the outcome value was imputed, and missed visit status). CDAI: Clinical Disease Activity Index; PtGA: patient's global assessment; PGA: physician's global assessment; CRP: C-reactive protein; RF: rheumatoid factor; DMARD: disease-modifying antirheumatic drugs; anti-CCP: anticyclic citrullinated peptide.

**Table 3.** Generalized estimating equation results demonstrating the expected difference in disease activity measures comparing smokers to nonsmokers over the study period.

Variables	$\beta$ (95% CI), p
CDAI	0.36 (−1.51 to 2.24), p = 0.71
PtGA	4.62 (−1.02 to 10.26), p = 0.11
PGA	−0.09 (−4.41 to 4.23), p = 0.97
Swollen joint count	0.08 (−0.73 to 0.90), p = 0.84
Tender joint count	0.42 (−0.35 to 1.18), p = 0.29
CRP, mg/l	−1.64 (−5.30 to 2.02), p = 0.38

An expected difference in disease activity for a unit change in the predictor (i.e., comparing smokers to nonsmokers) over the population, adjusted for baseline covariates (sex, age, race/ethnicity, past smoking, disease activity, obesity, RF status, anti-CCP status) and time-varying covariates (biologic/small molecule DMARD, synthetic DMARD, whether the outcome value was imputed, and missed visit status). CDAI: Clinical Disease Activity Index; PtGA: patient's global assessment; PGA: physician's global assessment; CRP: C-reactive protein; DMARD: disease-modifying antirheumatic drugs; RF: rheumatoid factor; anti-CCP: anticyclic citrullinated peptide.

Previous studies have shown strong evidence for an association between smoking and RA nodule formation<sup>3,4,5,6,7,8</sup>; however, additional measures of RA disease severity and activity are mixed. More radiographic damage was found in smokers compared to nonsmokers<sup>5</sup>, but not replicated in other studies<sup>9,10</sup>. Saag, *et al* found that erosions were associated with pack-years of smoking<sup>21</sup>, but a later study found a null association<sup>4</sup>. Some studies have demonstrated no difference in disease activity scores (28-joint count Disease Activity Score) between smokers and nonsmokers<sup>22,23,24</sup>. While other results showed that current smokers have higher CDAI scores than former or never smokers<sup>22,25</sup>, no significant differences were found between former and continuing smokers in a longitudinal study<sup>25</sup>. None of these studies incorporated methods to account for time-varying confounding or informative missingness.

Our findings show that interpretations surrounding the association between smoking and RA disease activity may differ dramatically depending on the type of statistical analysis conducted. We found no association between current smoking and any disease activity measure using conventional parametric methods (GEE). These findings are consistent with a longitudinal study that found no association between smoking and disease activity measures (SJC and TJC, and pain score) after 24 months<sup>9</sup>. Differences in results between conventional analytic methods and methods that properly adjusted for time-varying confounding, such as IPW, have been demonstrated in previous studies in other fields<sup>26,27</sup>. A review found substantial differences between causal models and conventional analyses in studies where time-varying confounding was suspected; IPW estimates differed from conventional estimates by at least 20% in about 40% of exposure-outcome associations in which the direction of the effect was the same<sup>27</sup>. Further, about 11%

of papers showed opposite interpretations between the 2 types of analyses.

An unexpected finding was that current smoking over time was inversely associated with TJC. This is in contrast to our finding that current smoking was significantly associated with a higher number of swollen joints. Discordance between the 2 measures has been previously described<sup>28</sup>. Michelsen, *et al* found that a patient's perspective on disease evaluation may differ from a physician's, which may affect disease activity measures such as TJC<sup>28</sup>. Authors attribute the difference to level of pain, which may be influenced by comorbid conditions such as fibromyalgia or psychosocial factors. Future studies are needed examining the association between smoking and TJC while adjusting for these additional factors.

We did not find an association between current smoking and CRP levels, in agreement with previous studies. While smoking has been associated with higher levels of CRP compared to nonsmokers in non-RA populations<sup>29</sup>, other studies have shown a null relationship<sup>30,31</sup>. One study in an RA population found that CRP levels were not associated with smoking status, and over time, CRP levels decreased in ever, never, and former smokers with no significant difference at any timepoint between groups<sup>9</sup>. CRP may be influenced by a number of factors independent of disease activity, including subtle synovitis not detected through clinical joint examinations, chronic inflammation at extra-articular sites, infection, and adipose tissue<sup>32</sup>.

Our study has several strengths, including a large, diverse population drawn from an integrated public hospital, longitudinal data, and robust methods that account for time-varying confounding and informative missingness. Limitations include the possibilities of uncontrolled bias, due to unmeasured confounding factors, inadequate control for missing data, and misspecification of exposure. Our approach to missing value imputation is based on the last value carried forward method, which, although relatively standard in longitudinal data analyses, could result in misclassified covariate values. We also note that social desirability bias is another possible source of misclassification of exposure data. We attempted to address the possibility of informative missingness and selection bias through adjustment for indicator variables for both missed visits and visits where imputations were made. However, there may be additional unmeasured time-dependent covariates that influence whether a patient was present in the clinic. In total, smoking was imputed for ~20% of all observations over the study period of 27 months.

As expected, there was loss to followup in this real-world clinical population over the study period. While we controlled for the potentially informative missing process with LTMLE, the smaller number of observed individuals may also contribute to a higher potential for violations of statistical assumptions. Continuous measure of obesity, rather than just baseline, may also improve estimation. Stratified

analyses based on variables such as RF and anti-CCP status, as well as race/ethnicity, were not feasible owing to small sample size and low power, which would have resulted in a higher probability of positivity violations using LTMLE. Additional studies are needed to examine how these variables may modify the association between smoking and RA disease activity. Lastly, we were unable to measure the effect of smoking on nodule formation or radiographic changes, or to examine associations beyond 27 months.

We rigorously examined the effect of current smoking status on disease activity using methods that account for time-varying confounding and potentially informative missingness. We found evidence that smoking is associated with higher levels of disease activity as measured by PtGA score and SJC. Methods used in our study may be useful for investigations of additional exposures on longitudinal outcome measures in rheumatologic disease.

## ONLINE SUPPLEMENT

Supplementary material accompanies the online version of this article.

## REFERENCES

1. Yelin E, Meenan R, Nevitt M, Epstein W. Work disability in rheumatoid arthritis: effects of disease, social, and work factors. *Ann Intern Med* 1980;93:551-6.
2. Sugiyama D, Nishimura K, Tamaki K, Tsuji G, Nakazawa T, Morinobu A, et al. Impact of smoking as a risk factor for developing rheumatoid arthritis: a meta-analysis of observational studies. *Ann Rheum Dis* 2010;69:70-81.
3. Wolfe F. The effect of smoking on clinical, laboratory, and radiographic status in rheumatoid arthritis. *J Rheumatol* 2000;27:630-7.
4. Mikuls TR, Hughes LB, Westfall AO, Holers VM, Parrish L, van der Heijde D, et al. Cigarette smoking, disease severity and autoantibody expression in African Americans with recent-onset rheumatoid arthritis. *Ann Rheum Dis* 2008;67:1529-34.
5. Matvey DL, Dawes PT, Fisher J, Brownfield A, Thomson W, Hajeer AH, et al. Nodular disease in rheumatoid arthritis: association with cigarette smoking and HLA-DRB1/TNF gene interaction. *J Rheumatol* 2002;29:2313-8.
6. Masdottir B, Jonsson T, Manfredsdottir V, Vikingsson A, Brekkan A, Valdimarsson H. Smoking, rheumatoid factor isotypes and severity of rheumatoid arthritis. *Rheumatology* 2000;39:1202-5.
7. Harrison BJ, Silman AJ, Wiles NJ, Scott DG, Symmons DP. The association of cigarette smoking with disease outcome in patients with early inflammatory polyarthritis. *Arthritis Rheum* 2001;44:323-30.
8. Papadopoulos NG, Alamanos Y, Voulgari PV, Epagelis EK, Tsifetaki N, Drosos AA. Does cigarette smoking influence disease expression, activity and severity in early rheumatoid arthritis patients? *Clin Exp Rheumatol* 2005;23:861-6.
9. Manfredsdottir VF, Vikingsdottir T, Jonsson T, Geirsson AJ, Kjartansson O, Heimisdottir M, et al. The effects of tobacco smoking and rheumatoid factor seropositivity on disease activity and joint damage in early rheumatoid arthritis. *Rheumatology* 2006;45:734-40.
10. Finckh A, Dehler S, Costenbader KH, Gabay C; Swiss Clinical Quality Management project for RA. Cigarette smoking and radiographic progression in rheumatoid arthritis. *Ann Rheum Dis* 2007;66:1066-71.
11. Robins JM, Hernan MA, Brumback B. Marginal structural models and causal inference in epidemiology. *Epidemiology* 2000; 11:550-60.
12. Robins JM. Analytic methods for estimating HIV treatment and cofactor effects. In: Ostrow DG, Kessler RC, editors. *Methodological issues of AIDS mental health research*. New York: Plenum Publishing; 1993:213-90.
13. Petersen M, Schwab J, Gruber S, Blaser N, Schomaker M, van der Laan M. Targeted maximum likelihood estimation for dynamic and static longitudinal marginal structural working models. *J Causal Inference* 2014;2:147-85.
14. Aletaha D, Nell VP, Stamm T, Uffmann M, Pflugbeil S, Machold K, et al. Acute phase reactants add little to composite disease activity indices for rheumatoid arthritis: validation of a clinical activity score. *Arthritis Res Ther* 2005;7:R796-806.
15. Robins JM. Marginal structural models versus structural nested models as tools for causal inference. In: *Statistical models in epidemiology, the environment, and clinical trials*. New York: Springer; 2000:95-133.
16. Hernan MA, Lanoy D, Costagliola D, Robins JM. Comparison of dynamic treatment regimes via inverse probability weighting. *Basic Clin Pharmacol Toxicol* 2006;98:237-42.
17. Petersen ML, Porter KE, Gruber S, Wang Y, van der Laan MJ. Diagnosing and responding to violations in the positivity assumption. *Stat Methods Med Res* 2012;21:31-54.
18. van der Laan MJ, Gruber S. Targeted minimum loss based estimation of causal effects of multiple time point interventions. *Int J Biostat* 2012;8.
19. Schwab J, Lendle S, Petersen M, van der Laan M, editors. *Ltmle: Longitudinal targeted maximum likelihood estimation*. R package version 0.9.3-1. [Internet. Accessed October 18, 2018.] Available from: [CRAN.R-project.org/package=ltmle](http://CRAN.R-project.org/package=ltmle)
20. Hanley JA, Negassa A, Edwards MD, Forrester JE. Statistical analysis of correlated data using generalized estimating equations: an orientation. *Am J Epidemiol* 2003;157:364-75.
21. Saag KG, Cerhan JR, Kolluri S, Ohashi K, Hunninghake GW, Schwartz DA. Cigarette smoking and rheumatoid arthritis severity. *Ann Rheum Dis* 1997;56:463-9.
22. Jacob B, Tomilionson G, Akahavan P, Bombardier C. Cigarette smoking and disease activity in rheumatoid arthritis patients: Results from Ontario Best Practice Research Initiative (OBRI) [abstract]. *Ann Rheum Dis* 2014;72 Suppl 2:329.
23. Fisher MC, Hochberg MC, El-Taha M, Kremer JM, Peng C, Greenberg JD; CORRONA Investigators. Smoking, smoking cessation, and disease activity in a large cohort of patients with rheumatoid arthritis. *J Rheumatol* 2012;39:904-9.
24. Westhoff G, Rau R, Zink A. Rheumatoid arthritis patients who smoke have a higher need for DMARDs and feel worse, but they do not have more joint damage than non-smokers of the same serological group. *Rheumatology* 2008;47:849-54.
25. Lu B, Rho YH, Cui J, Iannaccone CK, Frits ML, Karlson EW, et al. Associations of smoking and alcohol consumption with disease activity and functional status in rheumatoid arthritis. *J Rheumatol* 2014;41:24-30.
26. Li L, Evans E, Hser YI. A marginal structural modeling approach to assess the cumulative effect of drug treatment on the later drug use abstinence. *J Drug Issues* 2010;40:221-40.
27. Suarez D, Borrás R, Basagana X. Differences between marginal structural models and conventional models in their exposure effect estimates: a systematic review. *Epidemiology* 2011;22:586-8.
28. Michelsen B, Kristianslund EK, Hammer HB, Fagerli KM, Lie E, Wierod A, et al. Discordance between tender and swollen joint count as well as patient's and evaluator's global assessment may reduce likelihood of remission in patients with rheumatoid arthritis and psoriatic arthritis: data from the prospective multicentre NOR-DMARD study. *Ann Rheum Dis* 2017;76:708-11.

29. Ohsawa M, Okayama A, Nakamura M, Onoda T, Kato K, Itai K, et al. CRP levels are elevated in smokers but unrelated to the number of cigarettes and are decreased by long-term smoking cessation in male smokers. *Prev Med* 2005;41:651-6.
30. Jamal O, Aneni EC, Shaharyar S, Ali SS, Parris D, McEvoy JW, et al. Cigarette smoking worsens systemic inflammation in persons with metabolic syndrome. *Diabetol Metab Syndr* 2014;6:79.
31. Aldaham S, Foote JA, Chow HH, Hakim IA. Smoking status effect on inflammatory markers in a randomized trial of current and former heavy smokers. *Int J Inflam* 2015;2015:439396.
32. Graf J, Scherzer R, Grunfeld C, Imboden J. Levels of C-reactive protein associated with high and very high cardiovascular risk are prevalent in patients with rheumatoid arthritis. *PLoS One* 2009;4:e6242.