HHS Public Access

Author manuscript

Arthritis Rheumatol. Author manuscript; available in PMC 2018 August 01.

Published in final edited form as:

Arthritis Rheumatol. 2017 August ; 69(8): 1701–1702. doi:10.1002/art.40144.

Corrected estimates for the prevalence of self-reported doctor-diagnosed arthritis among US adults

S. Reza Jafarzadeh, DVM, MPVM, PhD[1] and **David T. Felson, MD, MPH**[1,2]

[1]Clinical Epidemiology Research and Training Unit, Boston University School of Medicine

[2]University of Manchester and Central Manchester NHS Foundation Trust, Manchester, UK

To the Editor

Hootman et al 2016 provided updated estimates for the current and future (i.e. projected) prevalence of self-reported doctor-diagnosed arthritis in the United States based on 2010–2012 National Health Interview Survey (NHIS) data (1). An individual with arthritis was identified in NHIS by an affirmation to the question "Have you ever been told by a doctor or other health professional that you have some form of arthritis, rheumatoid arthritis, gout, lupus or fibromyalgia?". The accuracy of this surveillance-aimed self-report case definition was previously reported in a validation study (2), and has been the basis for national estimates of arthritis prevalence. The validation study reported that the sensitivity and specificity of self-report doctor-diagnosed arthritis are 52.5% and 79.6% for ages 45–65 years, respectively, and 68.8% and 81.1% for those age ≥ 65 years, respectively.

Estimates for the prevalence of self-reported doctor-diagnosed arthritis are subject to misclassification bias, as these estimates are not adjusted for the imperfect (i.e. < 100%) and also variable accuracy (i.e. distinct sensitivities across age groups) of the case definition used (1). To estimate the *true* prevalence of arthritis in the United States after adjusting for this misclassification, we used a Bayesian approach, taking estimates for the *apparent* prevalence (i.e. the figure reported by Hootman et al), using the most recently-available 2015 NHIS data for adults from 18–64 years (we shall call this group under 65 years) and those 65 years and over (3). We formally incorporated the imperfect accuracy of the case definition as well as the uncertainty (i.e. variability) regarding these measures (e.g. sensitivities) in our analysis.

For Bayesian inference, we initially described past knowledge by specifying probability distributions (i.e. referred to as priors). We then updated the priors by observed data to obtain updated (i.e. posterior) estimates for unknown parameters (e.g. true prevalence) (4). For probabilities, such as sensitivity, specificity, or prevalence parameters, priors are commonly specified by *beta* distributions. A beta distribution with parameters a and b, $Beta(a, b)$, is a probability distribution that is confined to 0 and 1, where a and b define the shape of the distribution. For example, the mean and mode of beta distribution are given by

Address correspondence to S. Reza Jafarzadeh, DVM, MPVM, PhD, Clinical Epidemiology Research and Training Unit, Boston University School of Medicine, 650 Albany Street, Suite X200, Boston, MA 02118. Tel: 617-638-5884. Fax: 617-638-5239. srjafarz@bu.edu.

a/(a+b) and (a−1)/(a+b−2), respectively. Based on the validation study, we assumed that the most-likely value for the sensitivity of the case definition in the populations below and above 65 years of age was 52.5% and 68.8%, respectively. We further assumed that we are 95% certain that the sensitivity was less than 68.8% for the population below 65 years of age, and more than 52.5% for the population above 65 years of age. These two quantities are regarded as the mode and 5th- (or 95th-) percentile of beta distributions (5), and represented as *Beta*(19.04, 9.18) and *Beta*(12.30, 11.23) for the sensitivities of the case definition in the populations below and at or above 65 years of age, respectively. Even though not in the validation study, we further considered that the prevalence of arthritis in women was higher than that in men, implying a higher sensitivity of case definition in women of similar age groups. We specified *Beta*(48.28, 12.82) distribution for the specificity of the case definition across age strata, and non-informative prior *Beta*(1, 1) for the *true* prevalence of arthritis for all age and sex strata. The specified non-informative prior implied that every possible value between 0–100% for the true prevalence was equally likely.

Using data from the 2015 NHIS, 25.8% (8,689/33,672) responded *Yes* to the question on doctor-diagnosed arthritis. Bayesian estimates for the true prevalence of arthritis were 4.7% (95% probability interval [PI]: 0.2%, 17.8%) for men 18–64 years old, 16.1% (95% PI: 10.3%, 28.3%) for women 18–64 years old, 59.7% (95% PI: 47.8%, 76.9%) for men 65 or older, and 72.8% (95% PI: 59.7%, 92.9%) for women 65 or older. Estimates for the sensitivity of the case definition were 43.6% (95% PI: 27.2%, 58.7%) for men 18–64 years old, 53.9% (95% PI: 38.9%, 66.8%) for women 18–64 years old, 63.8% (95% PI: 52.7%, 75.6%) for men 65 or older, and 70.8% (95% PI: 58.5%, 83.0%) for women 65 or older. The specificity of the case definition was 86.3% (95% PI: 84.8%, 90.3%). Using these percentages and the projected US population from the US Census Bureau, we can estimate the current and projected burden of self-reported doctor-diagnosed arthritis among US adults (see Table 1).

Our results suggested that estimates provided by Hootman et al for the prevalence of doctor-diagnosed arthritis in the United States in 2015 were overestimated for men and underestimated for women. The imperfect specificity of the case definition also resulted in overestimation of the total prevalence since a substantial portion of non-diseased individuals were misclassified as having arthritis by Hootman et al. Our estimate for the projected *total* prevalence in 2040 was similar to that of Hootman et al, despite being substantially different when stratified by sex. This similarity occurred by chance, and was due to approximately similar numbers of individuals misclassified as diseased due to the imperfect specificity of the case definition, compared to those truly diseased individuals missed by imperfect sensitivity of the case definition in the total population in 2040. We suggest that future estimates of arthritis prevalence include an adjustment for imperfect case definition accuracy. The JAGS software package code that are used in the calculations can be supplied upon request.

Acknowledgments

Supported by the National Institutes of Health (grant AR-47785)

References

1. Hootman JM, Helmick CG, Barbour KE, Theis KA, Boring MA. Updated Projected Prevalence of Self-Reported Doctor-Diagnosed Arthritis and Arthritis-Attributable Activity Limitation Among US Adults, 2015–2040. Arthritis Rheumatol Hoboken NJ. 2016; 68:1582–1587.
2. Sacks JJ, Harrold LR, Helmick CG, Gurwitz JH, Emani S, Yood RA. Validation of a surveillance case definition for arthritis. J Rheumatol. 2005; 32:340–347. [PubMed: 15693097]
3. Messam LLM, Branscum AJ, Collins MT, Gardner IA. Frequentist and Bayesian approaches to prevalence estimation using examples from Johne's disease. Anim Health Res Rev. 2008; 9:1–23. [PubMed: 18346298]
4. Christensen, R., Johnson, WO., Branscum, AJ., Hanson, TE. Bayesian Ideas and Data Analysis: an Introduction for Scientists and Statisticians. 1. CRC Press; 2010.
5. Suess EA, Gardner IA, Johnson WO. Hierarchical Bayesian model for prevalence inferences and determination of a country's status for an animal pathogen. Prev Vet Med. 2002; 55:155–171. [PubMed: 12383652]

Table 1

Current and projected percent of the population with self-reported doctor-diagnosed arthritis among US adults, 2015–2060, adjusted for case definition imperfect accuracy

Year	Population Percentage with Arthritis (Projected Number of Individuals with Arthritis/Estimated Population)		
	Total	Men	Women
2015	21.4 (52.9/247.7)	14.4 (17.3/120.7)	28.0 (35.6/127.0)
2020	22.7 (59.0/260.4)	15.6 (19.9/127.2)	29.4 (39.1/133.2)
2025	24.1 (65.6/272.3)	17.0 (22.6/133.2)	30.9 (43.0/139.2)
2030	25.2 (71.3/283.1)	18.0 (24.9/138.5)	32.1 (46.4/144.6)
2035	25.7 (75.1/292.9)	18.4 (26.4/143.4)	32.6 (48.8/149.4)
2040	25.8 (77.8/302.0)	18.5 (27.3/148.1)	32.8 (50.5/153.9)
2045	25.8 (80.0/310.5)	18.5 (28.2/152.6)	32.8 (51.8/157.9)
2050	25.9 (82.6/318.4)	18.7 (29.3/156.9)	33.0 (53.3/161.5)
2055	26.3 (85.9/326.3)	19.2 (30.9/161.2)	33.3 (55.0/165.1)
2060	26.9 (89.9/334.5)	19.8 (32.8/165.7)	33.8 (57.0/168.8)

Values in parentheses are number of individuals in millions with arthritis divided the total population