Published in final edited form as:

Arthritis Rheumatol. 2019 January ; 71(1): 175–176. doi:10.1002/art.40732.

Adjustment for misclassification bias is necessary in survey-based arthritis surveillance to produce credible estimates of prevalence

S. Reza Jafarzadeh, DVM, MPVM, PhD and David T. Felson, MD, MPH

Clinical Epidemiology Research and Training Unit, Boston University School of Medicine, Boston, MA

We thank Murphy et al from the Centers for Disease Control and Prevention (CDC) for their comments on our work on arthritis prevalence. Murphy and colleagues insist that one survey question asking participants whether they recall an arthritis diagnosis by a health professional is sufficient to accurately characterize the prevalence of arthritis. Instead, we used three survey questions (theirs and two questions on joint pain) to increase the accuracy of arthritis prevalence estimation. Murphy et al asserted that in our paper, a positive response to “any” of the three arthritis-related survey questions in the National Health Interview Survey (NHIS) qualified as a case of arthritis, and this is incorrect.

Our approach starts with the assumption that, using a survey alone, it is not possible to accurately characterize a person’s arthritis status. When true disease status of individuals is unknown, latent class methods offer an unbiased estimation of prevalence in the population. The methods are based on calculating true- and false-positive fractions for each or a combination of survey responses. Given the imperfect sensitivity (Se) and specificity (Sp) of the questions used to screen for arthritis, the CDC approach which assumes that those who report doctor-diagnosed arthritis have arthritis is fundamentally flawed. In this situation, the only method to provide accurate prevalence estimates is one that adjusts for misclassification. We briefly describe how prevalence can be calculated from an imperfect screening instrument, e.g., a test, survey question, etc., followed by a note on using multiple survey questions simultaneously to estimate prevalence. We will point to some fundamental flaws in Murphy et al arguments, describe how our approach was misunderstood by authors, and conclude with comments and suggestions.

When the sensitivity (Se) and specificity (Sp) of an imperfect instrument (such as a single question to estimate arthritis prevalence) are known, an unbiased estimate for prevalence ($Prev$) can be calculated from apparent prevalence (i.e., proportion tested positive) by summing up true- and false-positive fractions (1), that is, $Pr(T+) = Prev Se + (1-Prev)(1-Sp)$. Besides Se and Sp , the only extra input required for unbiased estimation of prevalence is the proportion who tested positive, that is $Pr(T+)$. Unless a perfect instrument with $Se = Sp = 100\%$ is used, the proportion who tested positive produces a biased estimate of prevalence, and this bias can be substantial when Se and/or Sp is poor. Unfortunately, the survey

question (self-reported recall of doctor-diagnosed arthritis) used by the CDC to derive national estimates of arthritis prevalence has consistently been shown (2,3) to have poor Se and Sp , as exemplified by the validation study of Sacks et al (4) in which these questions were administered to a mixed group of patients who were then examined by trained rheumatology nurses asked to identify patients with treatable arthritis. Further, while this study reported that the Se for doctor-diagnosed arthritis question was merely 53% for persons ages 45–64, our own estimates (Table 3 in our paper) suggested Se of 22% and 34% in men and women (5), respectively, implying that using this single question for arthritis surveillance would miss nearly 65–80% of persons this age with arthritis.

An estimate of apparent prevalence based on the proportion tested positive by any screening instrument or survey question (the CDC's approach) should not be taken as an unbiased estimate of prevalence (6). We encourage future reports to use an adjusted estimate by, for example, applying the Rogan and Gladen's formulae (7) that accommodate true- and false-positive fractions.

Murphy and colleagues criticized the generalizability of the reported Se and Sp from Sacks et al for survey questions used in the NHIS. We agree with Murphy and colleagues that more comprehensive validation studies of arthritis-related survey questions are needed to provide assurance of surveillance instrument validity. However, we note that we used estimates of Se and Sp from the Sacks et al study only as guidance to construct prior probabilities for accuracy parameters (see our Supplementary Table 1). This means that we used this validation study to define a "distribution" of probable values for Se and Sp (8), rather than using the exact values provided in the study. This distinguishes our Bayesian approach from a frequentist method in estimating arthritis prevalence.

When three tests (e.g., survey questions) are available, true- and false-positive fractions for all realizations of test outcomes can be calculated just like the one-test scenario. For example, one realization of survey outcomes could be the proportion of participants who responded positively to all three arthritis-related NHIS questions, that is $\Pr(T_{1+}, T_{2+}, T_{3+})$, where T_{1+} is a positive response to the first survey question. Using corresponding Se and Sp , an unbiased estimate of prevalence can be derived from true- and false-positive fractions through $\Pr(T_{1+}, T_{2+}, T_{3+}) = Prev Se_1 Se_2 Se_3 + (1-Prev)(1-Sp_1)(1-Sp_2)(1-Sp_3)$. As noted in our paper, the observed frequency of each possible realization is modeled as a multinomial sampling distribution without the need to know the true disease status of each participant (9). Both models (i.e., the one with one test and the other with three tests) target the same parameter, i.e. arthritis prevalence. We referred to this multiple testing technique as an expanded "surveillance definition" in our paper.

While we cannot know the arthritis status of each individual from survey questions, it is possible to provide an unbiased estimate of prevalence in the population by using all realizations of responses to survey questions along with their corresponding Se and Sp .

If arthritis prevalence can be accurately estimated using one question adjusted for misclassification, why are three questions about arthritis and joint pain preferable? There are two reasons. First, the misclassification of arthritis cases is less when multiple overlapping

questions about arthritis and its symptoms are used so that adjusting for misclassification produces a more accurate assessment of arthritis prevalence.

Second, we are concerned about another public health consequence of the CDC's approach that identifies as having arthritis only those who recall having been seen by a health practitioner and given an arthritis diagnosis. At a time when many in our society remain outside the medical care system, this approach will make it impossible to know whether substantial numbers of persons are suffering from undiagnosed arthritis.

Lastly, Murphy et al opposed our suggestion to improve NHIS survey questions by including osteoarthritis, the most common form of arthritis, and excluding fibromyalgia from the doctor-diagnosed arthritis question. We maintain our belief that NHIS survey questions aim to estimate the burden of arthritis and should not include questions about non-arthritis conditions. Regardless, the impact of this shortcoming can be minimized if techniques to adjust for misclassification bias are implemented.

In conclusion, we note that reporting the proportion who responded positively to one survey question on doctor-diagnosed arthritis does not produce accurate national estimates of arthritis prevalence and the consequence, as we noted, has been to grossly underestimate its prevalence. Further, regardless of how many and which survey instruments are used, providing estimates that are adjusted for misclassification bias is essential to produce credible estimates, estimates that could guide efficient planning at federal and local levels and could be used to measure the arthritis burden.

References

1. Messam LLM, Branscum AJ, Collins MT, Gardner IA. Frequentist and Bayesian approaches to prevalence estimation using examples from John's disease. *Anim Health Res Rev* 2008;9:1–23. [PubMed: 18346298]
2. Bombard JM, Powell KE, Martin LM, Helmick CG, Wilson WH. Validity and reliability of self-reported arthritis: Georgia senior centers, 2000–2001. *Am J Prev Med* 2005;28:251–258. [PubMed: 15766612]
3. Lo T, Parkinson L, Cunich M, Byles J. Discordance between self-reported arthritis and musculoskeletal signs and symptoms in older women. *BMC Musculoskelet Disord* 2016;17:494. [PubMed: 27905906]
4. 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]
5. Jafarzadeh SR, Felson DT. Updated estimates suggest a much higher prevalence of arthritis in United States adults than previous ones. *Arthritis Rheumatol* Hoboken NJ 2018;70:185–192.
6. Enøe C, Georgiadis MP, Johnson WO. Estimation of sensitivity and specificity of diagnostic tests and disease prevalence when the true disease state is unknown. *Prev Vet Med* 2000;45:61–81. [PubMed: 10802334]
7. Rogan WJ, Gladen B. Estimating prevalence from the results of a screening test. *Am J Epidemiol* 1978;107:71–76. [PubMed: 623091]
8. Branscum AJ, Gardner IA, Johnson WO. Estimation of diagnostic-test sensitivity and specificity through Bayesian modeling. *Prev Vet Med* 2005;68:145–163. [PubMed: 15820113]
9. Collins J, Huynh M. Estimation of diagnostic test accuracy without full verification: a review of latent class methods. *Stat Med* 2014;33:4141–4169. [PubMed: 24910172]