



HHS Public Access

Author manuscript

Arthritis Rheumatol. Author manuscript; available in PMC 2019 February 01.

Published in final edited form as:

Arthritis Rheumatol. 2018 February ; 70(2): 185–192. doi:10.1002/art.40355.

Updated estimates suggest a much higher prevalence of arthritis in US adults than previous ones

S. Reza Jafarzadeh, DVM, MPVM, PhD¹ and David T. Felson, MD, MPH^{1,2}

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

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

Abstract

Objective—National estimates of arthritis prevalence relies on a single survey question about doctor-diagnosed arthritis without using survey information on joint symptoms, even though some subjects with only the latter have been shown to have arthritis. The sensitivity of the current surveillance definition is only 53% and 69% in subjects 45–64 and 65 years of age, respectively, resulting in misclassification of nearly half and one-third of subjects in those age groups. We aimed to estimate arthritis prevalence based on an expansive surveillance definition, that is also adjusted for the measurement errors in the current definition.

Methods—Using the 2015 National Health Interview Survey, we developed a Bayesian multinomial latent class model for arthritis surveillance based on doctor-diagnosed arthritis, joint symptoms, and whether symptom duration exceeded three months.

Results—Of 33,672 participants, 19.3% of men and 16.7% of women age 18–64 affirmed joint symptoms without doctor-diagnosed arthritis; proportions were 15.7% and 13.5%, respectively, for those 65. The measurement error-adjusted prevalence of arthritis was 29.9% (95% Bayesian probability interval [PI]: 23.4, 42.3) in men 18–64, 31.2% (95% PI: 25.8, 44.1) in women 18–64, 55.8% (95% PI: 49.9, 70.4) in men 65, and 68.7% (95% PI: 62.1, 79.9) in women 65. Arthritis affected 91.2 (of 247.7; 36.8%) million adults in the US in 2015, which included 61.1 (of 199.9; 30.6%) million persons between 18–64. Our prevalence estimate is 68% higher than previously reported arthritis national estimate.

Conclusion—Arthritis prevalence in the US population has been substantially underestimated, especially among adults <65 years.

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.

DR. S. REZA JAFARZADEH (Orcid ID: 0000-0002-1099-9175)

AUTHOR CONTRIBUTIONS

All authors were involved in drafting the article or revising it critically for important intellectual content, and all authors approved the final version to be published.

Study conception and design. Jafarzadeh, Felson.

Acquisition of data. Jafarzadeh.

Analysis and interpretation of data. Jafarzadeh, Felson

INTRODUCTION

Arthritis is a highly prevalent condition in the United States and a leading cause of disability. The economic burden of arthritis is estimated to be at least \$128 billion annually in the United States (1). Effective surveillance of arthritis on a national scale is challenging and requires a screening strategy that goes beyond recognizing symptoms reported in a clinical setting.

National surveillance efforts for arthritis rely on self-report surveys as a practical tool to estimate the burden of disease. The Centers for Disease Control and Prevention (CDC) routinely publishes estimates for the prevalence of arthritis in the United States (2–4). One source of data used for arthritis surveillance is the National Health Interview Survey (NHIS), administered by the U.S. Census Bureau, which includes questions that are used to identify cases of arthritis. Although identifying subjects with arthritis from these health surveys is a reasonable method for national surveillance efforts, the accuracy of estimates depends on the validity of the surveillance definition to identify cases with arthritis. The main question used from this survey to identify cases of arthritis has been a single question asking subjects if they have doctor-diagnosed arthritis.

In a validation study in which they actually verified clinical cases of arthritis, Sacks et al (5) documented the diagnostic sensitivity and specificity of the arthritis-related survey questions. Survey validation has shown reassuring but imperfect accuracy. While a survey approach using a report of doctor-diagnosed arthritis had a higher sensitivity (68.8%) among those 65 and older, the sensitivity of this surveillance definition was lower (52.5%) for persons 45–64 years of age. Such a low sensitivity, especially in a younger population, where almost half of true arthritis cases are missed, results in substantial misclassification and underestimation of prevalence, and would have a detrimental effect for planning and needs assessment (3,4).

Since 2002, national estimates for the prevalence of arthritis or of doctor-diagnosed arthritis, that relied on the assurance of the Sacks et al (5) validation study, have produced an uncorrected estimate of 54.4 million adults (22.7%) in the United States in 2015 (3,4). No figures have been released that correct these estimates for the measurement errors caused by the imperfect sensitivity and specificity of surveillance definitions (3,4,6). Further, this likely underestimation of arthritis prevalence, especially in subjects age 45–64, has suggested that prevalence in this age group is low at a time when other reports noted marked increase in the rates of knee and hip replacement in this age group (7).

Strategies exist to increase the accuracy of surveillance criteria, such as combining the results of multiple individual diagnostic criteria. For example, one diagnostic criterion could be based on a person self-reporting a diagnosis of arthritis from a health professional. Another diagnostic criterion could be a person self-reporting symptoms that are consistent with arthritis. Questions about chronic joint symptoms are in fact included in the NHIS, and the Sacks et al (5) validation study reported that some subjects with chronic joint symptoms, without a doctor-diagnosed arthritis report, had a clinical diagnoses of arthritis. Nonetheless, chronic joint symptoms have not been used in combination to doctor-diagnosed arthritis to

derive national estimates of arthritis prevalence. While self-reported doctor-diagnosed arthritis have an acceptable specificity (i.e. %81.1) for arthritis in adults age 65 and over (5), many persons below 65 years of age did not report a diagnosis from a health professional despite reporting chronic joint symptoms.

In this work, we develop a Bayesian model to estimate the prevalence of arthritis among adults in the United States in 2015, that is, an estimate adjusted for the measurement errors due to the imperfect accuracy of surveillance criteria based on both the report of chronic joint symptoms and doctor-diagnosed arthritis. We used the term adjusted prevalence, in contrast to unadjusted prevalence, for our measurement error-corrected estimates for prevalence (8–13). We note that an estimate for the adjusted prevalence from a survey is not equivalent to the exact number of rheumatologist-verified arthritis cases, even though the survey questions were actually validated against such cases, but we use the word adjusted to suggest that we are correcting estimates for the systematic underestimation of prevalence that occurs when surveillance instruments with imperfect sensitivity are used.

METHODS

Study setting and data

We obtained the most recent publicly-available Sample Adult Core from the 2015 NHIS data release that contains data for individuals 18 years of age or older. NHIS, which is routinely used to derive national estimates of arthritis prevalence, is one of the most prominent population health surveys that covers the noninstitutionalized population in the United States; it excludes long-term care facilities, active duty armed forces personnel, or U.S. nationals living in a foreign country.

As noted in Sacks et al (5), the National Arthritis Data Workgroup suggested “arthritis” to be broadly defined as a condition with clinical significance that is either symptomatic or require attention from a health professional for treatment. The purpose of the definition, which excluded injuries, was to have a practical method to estimate the burden and impact of arthritis. For example, a case of asymptomatic radiographic osteoarthritis as a result of a previous injury was not considered clinically significant nor were asymptomatic Heberden’s nodes.

In our study, identical to the definition used by the CDC, a case of doctor-diagnosed arthritis was defined as a positive response to the following NHIS survey 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?”. In addition to doctor-diagnosed arthritis, a separate set of questions inquired into chronic joint symptoms in the NHIS, defined as a positive response to the question, “The next questions refer to your joints. Please do not include the back or neck. During the past 30 days, have you had any symptoms of pain, aching, or stiffness in or around a joint?”. Moreover, there was a follow-up question on the onset of symptoms, if the person reported recent chronic joint symptoms, as follow: “Did your joint symptoms first begin more than 3 months ago?”.

We developed surveillance criteria based on the three questions that were used to define doctor-diagnosed arthritis, chronic joint symptoms and whether the symptoms onset exceeded 3-month.

Surveillance criteria

We considered each of the three questions, described in the previous section, as a diagnostic test with imperfect accuracy for arthritis. The third question on the onset of symptoms was only possible (i.e. positive or negative) if the person reported the existence of recent chronic joint symptoms. Therefore, the data consisted of frequencies corresponding to one of the six possible realizations of test outcomes, (+/+/, +/-, -/+/, -/-, -/+/-), for doctor-diagnosed arthritis, chronic joint symptoms, and duration of symptoms, respectively. The null value (.) indicates that the value for the duration of symptoms was not available due to a negative response for recent chronic joint symptoms criterion. We further stratified the results of the surveillance criteria into 4 sub-populations based on sex and age groups 18–64 years and 65 years and over (Table 1).

Model

We developed a Bayesian multinomial latent class model for the 6 realizations of test outcomes in the 4 sub-populations presented in Table 1. Bayesian latent class models have been previously used in variety of models for diagnostic test outcomes when a perfect reference standard is not available (8–21). Latent class models do not require the true disease status of each subject to be known (i.e. observed) in order to estimate prevalence and measures of diagnostic accuracy (12,13). The multinomial probabilities corresponding to the observed frequencies of surveillance criteria were defined as functions of true prevalence and the sensitivity and specificity of each criterion, as described in Branscum et al (9) and others (10). For example, the probability of observing (+/+/) frequency is a product of true arthritis prevalence and the sensitivities of the three criteria in the surveillance definition, which is the true positive fraction, plus the product of truly non-arthritis population prevalence (i.e. one minus true arthritis prevalence) and the false positive fraction, which is given by one minus the specificities of the three criteria. All multinomial probabilities corresponding to observed frequencies are enumerated as presented in the Supplementary Materials.

As shown in the validation study by Sacks et al (5) and in other studies (17,22), the sensitivity of a diagnostic test is often higher when applied to a population with higher prevalence (in this case, this would be true of older vs. younger subjects). This occurs in part because there tends to be more severe disease in a high-prevalence population (17,23). In general, the diagnostic specificity, the probability of a negative outcome in a truly healthy (i.e. non-diseased) population, where the prevalence remains constant at zero, is less variable across non-diseased populations. To obtain a more robust estimate for arthritis prevalence (15,17), we allowed the sensitivity of the surveillance criteria to differ across the 4 sub-populations of men 18–64 years old, women 18–64 years old, men 65 years of age or older, and women 65 years of age or older. In an alternative parameterization for the purpose of sensitivity analysis, we assumed the sensitivity of the surveillance criteria to be the same among men and women, per estimates by Sacks et al (5), but to be different only by age (i.e.

higher sensitivity for older population). The alternative parameterization involves less parameters to be estimated with the same degrees of freedom (i.e. 2 sensitivities for each criterion instead of 4 sensitivities for each criterion in the primary model). Number of parameters and degrees of freedom affect model identifiability as we discuss in the next section.

The diagnostic specificity of criteria can be increased by serial interpretation of individual criteria results, that is, criteria considered positive when all individual components are positive. On the other hand, parallel interpretation of diagnostic criteria, that is, criteria considered positive when any individual component is positive, results in increased diagnostic sensitivity at the expense of reduced specificity. Similarly, sequential interpretation of criteria, where an individual criterion result is available only if another criterion is positive or negative, could result in improved sensitivity or specificity. Since subjects with symptoms such as pain are more likely to seek a health professional and receive a diagnosis for arthritis, we included conditional covariances, as described in Dendukuri and Joseph (24), to account for the potential dependence between the outcomes for doctor-diagnosed arthritis and chronic joint symptoms. Conditional dependence affects the joint-testing sensitivity and specificity because the sensitivity (or specificity) of a test would not be independent of the outcome of another test (25). A positive dependence between the sensitivities of the two tests occurs when the sensitivity of a test is lower among truly diseased subjects that are test-negative on the other test and vice versa. Consequently, a positive or negative dependence between the sensitivities of two diagnostic criteria increases or decreases, respectively, the gain in serial joint-testing sensitivity. Similarly, a dependence between tests specificities affect joint-testing accuracy (25).

Bayesian inference and priors

We used a Bayesian approach (26) to estimate the parameters of the multinomial latent class model for cross-classified outcomes of the arthritis surveillance criteria. In this approach, probability distributions are specified for model parameters, which consisted of the arthritis prevalences for the 4 sub-populations, the sensitivities and specificities of the surveillance criteria, and the conditional covariances between the outcomes of doctor-diagnosed arthritis and chronic joint symptoms criteria. These probability distributions are referred to as priors and are elicited from past knowledge or expert opinion, or specified to be non-informative when every possible value of the parameter is defined to have an equal probability of occurring. The prior distributions are updated with the observed data to obtain posterior distributions for the parameters of the model using Markov Chain Monte Carlo techniques (26). The Monte Carlo-based posterior distributions are then summarized as mean, median, or mode and the 2.5th and 97.5th percentiles of the Monte Carlo samples as Bayesian 95% probability intervals. For a description of prior elicitation and all the priors specified for the parameters of the multinomial model, see Supplementary Materials including Supplementary Table 1.

In latent class models, non-identifiability occurs when the model cannot guarantee a unique set of parameter estimates, often due to insufficient degrees of freedom (27). Non-identifiability can be mitigated with proper informative priors or putting constraints on

priors in Bayesian analysis (15,27,28). Hence, we ordered priors on prevalences and sensitivities in the sub-populations such that the prior distribution mean was higher in the older population than younger and higher in women than men.

Bayesian analysis was performed in JAGS software (29) version 4.2.0 through rjags package (30) version 4–6 in R software (31) version 3.3.3. Beta priors were elicited using epiR package (32) version 0.9–79 in R. The program code for running Bayesian computations was adapted from Branscum et al (9).

RESULTS

Table 1 presents the cross-classified outcomes of the arthritis surveillance criteria for 33,672 participants in 2015 NHIS. In subjects 18–64 years of age, 19.3% (2,242/11,597) of men and 16.7% (2,294/13,697) of women responded “yes” to the question on chronic joint symptoms, regardless of whether symptoms duration exceeded 3 months, but responded “no” to the doctor-diagnosed question (Table 1). For those 65 years of age or older, 15.7% (545/3,474) of men and 13.5% (660/4,904) of women responded “yes” to chronic joint symptoms, regardless of symptoms onset, without a concurrent report of doctor-diagnosed arthritis (Table 1).

The proportion who responded “yes” to doctor-diagnosed arthritis, with or without a concurrent report of chronic joint symptoms or symptoms onset if any, was 15.0% (1,740/11,597) for men 18–64 years of age, 20.0% (2,734/13,697) in women 18–64 years old, 43.5% (1,511/3,474) for men 65 or older, and 55.1% (2,704/4,904) in women 65 years of age or older (Table 1).

Posterior estimates and the corresponding 95% probability intervals (PI) for the measurement error-adjusted prevalences in the 4 sub-populations stratified by age and sex are presented in Table 2. The posterior median for the adjusted prevalence of arthritis based on the primary model was 29.9% (95% PI: 23.4%, 42.3%) in men 18–64 years of age, 31.2% (95% PI: 25.8%, 44.1%) in women 18–64 years of age, 55.8% (95% PI: 49.9%, 70.4%) in men 65 years old or older, and 68.7% (95% PI: 62.1%, 79.9%) in women 65 years old or older. The results of the sensitivity analysis that used identical values for sensitivity of the criteria in men and women, suggested similar estimates (i.e. with overlapping probability intervals) to the results of the primary analysis (Table 2).

The accuracy of the surveillance criteria is provided in Table 3. The results suggested very low sensitivity for doctor-diagnosed arthritis criterion in subjects 18–64 years of age, and the highest sensitivity for symptoms onset criterion across all age and sex strata, despite having the lowest specificity. Thus, a substantial (i.e. 65–80%) fraction of the population with arthritis, who are between 18–64 years of age, but are misclassified as healthy by the doctor-diagnosed arthritis criterion due to low sensitivity, are captured by the two remaining questions on joint pain, aching or stiffness.

Finally, the estimated number of adults with arthritis in the United States, based on the 2015 National Population Projections provided by the U.S. Census Bureau (33), was 91.2 million individuals (of the 247.7 million total population projection; 36.8%) that included 29.8

million men 18–64 years of age, 11.8 million men 65 or older, 31.3 million women 18–64 years of age, and 18.3 million women 65 years of age or older.

DISCUSSION

Using NHIS data, we developed an arthritis surveillance definition to estimate the measurement error-adjusted prevalence of arthritis in the United States, based on three criteria with imperfect accuracy, i.e. doctor-diagnosed arthritis, chronic joint symptoms, and symptom that had been present longer than 3 months. Our estimate suggested that 91.2 million adults (36.8%) in the United States were affected by arthritis in 2015. Our results suggested that the adjusted prevalence of arthritis, doctor-diagnosed or otherwise, is substantially higher than a previously-reported uncorrected estimate for the prevalence of doctor-diagnosed arthritis of 54.4 million adults (22.7%) in the United States (4) and also higher than the estimate for the adjusted prevalence of doctor-diagnosed arthritis of 52.9 million adults (21.4%) (34). Further, we estimated that 61.1 million adults between the ages of 18–64 (of 199.9 million total adults 18–64 years of age; 30.6%) had arthritis in the United States in 2015.

The higher prevalence that we report is due in large part to the previous underestimate of arthritis in adults between 18–64 years of age. Recent reports have suggested a marked increase in total knee replacement utilization, especially in the population of 45–64 years of age, that has outpaced the increasing rate of obesity in the same age group (7). Another study has reported higher arthritis prevalence in more recent birth cohorts, compared to previous generations of the same age, partly due to changing patterns of obesity in relatively younger populations (35). Individuals under the age of 65 may perceive arthritis as a condition affecting only the elderly; thus, may visit a health professional less often, or may ignore occasional joint symptoms. Moreover, arthritis may not be reported on electronic health records or insurance claims data if arthritis is not the primary reason for a referral to a healthcare provider. A previous study reported that of the total of 13.7% (6,064/44,326) adults in 2005 NHIS data who had chronic joint symptoms, but no indication of doctor-diagnosed arthritis, 89.1% were below 65 years of age (36), compared to 79.0% (4,536/5,741) in our study population in 2015 (Table 1).

Previous studies that supported (37,38) or contrasted (36) the addition of chronic joint symptoms to arthritis surveillance definition, through creating a pseudo-gold standard based on other criteria such as functional or activity limiting factors or any other indication of arthritis, are subject to the same flaws and limitations of relying on an imperfect surveillance definition. In contrast, our latent class analytic approach did not rely on the assumption of having a perfect reference standard (i.e. a gold standard) and did not require us to identify the true arthritis status of each individual in the population in order to estimate the measurement error-adjusted prevalence (8,39).

The question on doctor-diagnosed arthritis in the NHIS includes fibromyalgia among the conditions under the arthritis rubric. While fibromyalgia can cause joint pain and lead to a diagnosis from a health professional, it is not a form of arthritis. Consequently, inclusion of fibromyalgia resulted in an imperfect specificity for doctor-diagnosed arthritis criterion,

which affected the uncorrected estimates for the prevalence reported in previous studies (2–4); however, this inclusion did not affect our estimates for the adjusted prevalence because we already corrected our estimate for the imperfect specificity of the doctor-diagnosed arthritis criterion. Conversely, absence of osteoarthritis in the NHIS question on doctor-diagnosed arthritis, the most prevalent form of arthritis, results in an imperfect sensitivity for doctor-diagnosed arthritis criterion and subsequently affected uncorrected estimates for the prevalence in previous reports (2–4). In addition to problems with measurement errors, there are shortcomings in reliance of the previously published (3,4) national estimates of arthritis prevalence on a single survey question of doctor-diagnosed arthritis. Implicit in the question on doctor-diagnosed arthritis, when the response is positive, is that the surveyed individual sought or had access to medical care from a health professional. However, a negative response to the doctor-diagnosed question could be the result of either lack of medical attention to joint symptoms or a truly negative diagnosis with regard to arthritis. Moreover, an individual who is diagnosed by a health professional to have arthritis, may never be explicitly informed of the diagnosis.

The chronic joint symptoms question does not require pain on more than half of days and could represent mild or moderate joint pain. We note that the Sacks et al (5) validation study reported the sensitivity and specificity of these questions for subjects aged 45–64 years of age. We generalized these estimates to those aged 18–64 years, and it is conceivable that our arthritis prevalence estimates for persons aged 18–44 are off if these estimates of sensitivity and specificity are imprecise. However, our Bayesian approach mitigated this potential inaccuracy by specifying diffuse prior distributions that covered a wide range of sensitivity and specificity values, in contrast to frequentist approach where these values are assumed to be fixed.

The NHIS has a complex survey sample design, and our approach did not use a weighting scheme to estimate prevalence. There are currently two conceptually-competing and fundamentally distinct approaches to make “inferences” from complex surveys, the classical design-based (randomization-based) approach that uses a weighting scheme, described in Neyman (40), and our model-based approach that relies on developing statistical models to infer population parameters. While weights are useful for “designing” a cost-effective survey sample, their use in inference after survey data are collected is debated in statistical literature because relying on weighting alone may fail to sufficiently account for other factors that influence the accuracy of estimation such as response rate or misclassification. Weights are not attributes of the individuals or a particular disease under study, but are constructed as a product of probability calculations to correct for the perceived differences between a sample and a target population based on “design” variables such as age, sex, or location (41). There is a large body of statistical literature on the philosophical and fundamental differences between the design-based and the model-based approaches (for example see references 41–43). Some argue that reliance on a general framework to calculate weights, which requires many arbitrary choices on weighting factor, pooling, or truncation of weights, does not provide much benefit over using a model-based approach to directly estimate parameters of interest, especially when auxiliary data are available (i.e. accuracy of survey questions) or when faced with biases unrelated to sampling weights such as measurement error. While the federal agencies has historically produced statistical summaries using the design-based

approach (44), the U.S. Census Bureau recently formed a Research and Methodology Directorate to more effectively utilize model-based approaches for inference in official statistics by federal agencies (43). Our approach was based on a model to adjust for the sensitivities and specificities of the test's performance in the 4 sub-populations identified by sex and age. We acknowledge that the inability to directly apply weighting in our model-based approach may have introduced inaccuracy with regard to the precision of our estimates (i.e. wider probability intervals for the probability of true population parameters), but the gain in accuracy with regard to misclassification bias is so substantial that we believe it justifies our choice.

The validity of our Bayesian inference relies on the correct specification and estimation of the underlying probability distributions that generated the observed frequencies of the NHIS questions' outcomes in the 4 sub-populations shown in Table 1. Our modeling approach does not specify distinct priors on the diagnostic sensitivity and specificity of the NHIS questions across states, because the validation study of Sacks et al did not provide evidence that the accuracy of NHIS questions varied by state. Therefore, regardless of the sampling unit from which an individual was selected, for example Massachusetts versus in New York, the probability of observing a specific realization of outcomes for the NHIS questions for an individual relies only on whether the individual has arthritis, and the sensitivities and specificities of the questions.

Our model-based approach provides several advantages over previous studies that did not correct for measurement errors (2–4). In addition to directly applying the sensitivity and specificity of NHIS survey questions to our estimates, we allowed the correlation (i.e. conditional dependence) between responses to survey questions to be formally incorporated to get corrected estimates. Further, the Bayesian approach provided a coherent framework for model-based re-validation (i.e. that does not rely on a gold standard) of the findings of the Sacks et al (5) study, through updating the sensitivities and specificities of the survey questions from the validation study with the 2015 NHIS data, to obtain posterior probabilities (i.e. re-validated estimates), which we presented in Table 3.

The underlying rheumatic diseases resulting in arthritis are diverse. While our inference was limited to aggregate-level population surveillance on the burden of arthritis, further studies are needed to evaluate potential changes in the specific causes of arthritis, especially among adults below the age of 65. Arthritis causes an enormous economic and public health implications. Arthritis-attributable healthcare direct costs or long-term indirect costs as a result of loss of productivity and disability need be revised to account for the corrected prevalence of arthritis affecting individuals at younger age than previously perceived (1,45).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

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

References

1. Cisternas MG, Murphy LB, Yelin EH, Foreman AJ, Pasta DJ, Helmick CG. Trends in medical care expenditures of US adults with arthritis and other rheumatic conditions 1997 to 2005. *J Rheumatol*. 2009; 36:2531–2538. [PubMed: 19797505]
2. Hootman JM, Helmick CG. Projections of US prevalence of arthritis and associated activity limitations. *Arthritis Rheum*. 2006; 54:226–229. [PubMed: 16385518]
3. 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.
4. Barbour KE, Helmick CG, Boring M, Brady TJ. Vital signs: prevalence of doctor-diagnosed arthritis and arthritis-attributable activity limitation - United States, 2013–2015. *MMWR Morb Mortal Wkly Rep*. 2017; 66:246–253. [PubMed: 28278145]
5. 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]
6. Murphy LB, Cisternas MG, Greenlund KJ, Giles W, Hannan C, Helmick CG. Defining arthritis for public health surveillance: methods and estimates in four US population health surveys. *Arthritis Care Res*. 2017; 69:356–367.
7. Losina E, Thornhill TS, Rome BN, Wright J, Katz JN. The dramatic increase in total knee replacement utilization rates in the United States cannot be fully explained by growth in population size and the obesity epidemic. *J Bone Joint Surg Am*. 2012; 94:201–207. [PubMed: 22298051]
8. Branscum AJ, Gardner IA, Johnson WO. Bayesian modeling of animal- and herd-level prevalences. *Prev Vet Med*. 2004; 66:101–112. [PubMed: 15579338]
9. 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]
10. Ladouceur M, Rahme E, Pineau CA, Joseph L. Robustness of prevalence estimates derived from misclassified data from administrative databases. *Biometrics*. 2007; 63:272–279. [PubMed: 17447953]
11. 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]
12. 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]
13. Kostoulas P, Nielsen SS, Branscum AJ, Johnson WO, Dendukuri N, Dhand NK, et al. STARD-BLCM: Standards for the Reporting of Diagnostic accuracy studies that use Bayesian Latent Class Models. *Prev Vet Med*. 2017; 138:37–47. [PubMed: 28237234]
14. Joseph L, Gyorkos TW, Coupal L. Bayesian estimation of disease prevalence and the parameters of diagnostic tests in the absence of a gold standard. *Am J Epidemiol*. 1995; 141:263–272. [PubMed: 7840100]
15. Johnson WO, Gastwirth JL, Pearson LM. Screening without a “gold standard”: the Hui-Walter paradigm revisited. *Am J Epidemiol*. 2001; 153:921–924. [PubMed: 11323324]
16. Berkvens D, Speybroeck N, Praet N, Adel A, Lesaffre E. Estimating disease prevalence in a Bayesian framework using probabilistic constraints. *Epidemiol Camb Mass*. 2006; 17:145–153.
17. Johnson WO, Gardner IA, Metoyer CN, Branscum AJ. On the interpretation of test sensitivity in the two-test two-population problem: assumptions matter. *Prev Vet Med*. 2009; 91:116–121. [PubMed: 19651450]
18. Jafarzadeh SR, Johnson WO, Utts JM, Gardner IA. Bayesian estimation of the receiver operating characteristic curve for a diagnostic test with a limit of detection in the absence of a gold standard. *Stat Med*. 2010; 29:2090–2106. [PubMed: 20603894]
19. Jafarzadeh SR, Warren DK, Nickel KB, Wallace AE, Mines D, Fraser VJ, et al. Bayesian estimation of the accuracy of ICD-9-CM- and CPT-4-based algorithms to identify cholecystectomy procedures in administrative data without a reference standard. *Ann Epidemiol*. 2013; 23:592.

20. Jafarzadeh SR, Johnson WO, Gardner IA. Bayesian modeling and inference for diagnostic accuracy and probability of disease based on multiple diagnostic biomarkers with and without a perfect reference standard. *Stat Med.* 2016; 35:859–876. [PubMed: 26415924]
21. Jafarzadeh SR, Thomas BS, Gill J, Fraser VJ, Marschall J, Warren DK. Sepsis surveillance from administrative data in the absence of a perfect verification. *Ann Epidemiol.* 2016; 26:717–722.e1. [PubMed: 27600804]
22. Jafarzadeh, SR. Bayesian Methods For Evaluation Of Diagnostic Accuracy Of Quantitative Tests And Disease Diagnosis In The Absence Of A Perfect Reference Standard With Examples From Johne's Disease. 2012. Available at: <http://gradworks.umi.com/35/44/3544743.html>. Accessed April 19, 2013
23. Greiner M, Gardner IA. Epidemiologic issues in the validation of veterinary diagnostic tests. *Prev Vet Med.* 2000; 45:3–22. [PubMed: 10802331]
24. Dendukuri N, Joseph L. Bayesian approaches to modeling the conditional dependence between multiple diagnostic tests. *Biometrics.* 2001; 57:158–167. [PubMed: 11252592]
25. Gardner IA, Stryhn H, Lind P, Collins MT. Conditional dependence between tests affects the diagnosis and surveillance of animal diseases. *Prev Vet Med.* 2000; 45:107–122. [PubMed: 10802336]
26. Christensen, R., Johnson, WO., Branscum, AJ., Hanson, TE. *Bayesian Ideas and Data Analysis: an Introduction for Scientists and Statisticians.* 1st. CRC Press; 2010.
27. Jones G, Johnson WO, Hanson TE, Christensen R. Identifiability of models for multiple diagnostic testing in the absence of a gold standard. *Biometrics.* 2010; 66:855–863. [PubMed: 19764953]
28. Georgiadis MP, Johnson WO, Gardner IA, Singh R. Correlation-adjusted estimation of sensitivity and specificity of two diagnostic tests. *J R Stat Soc Ser C Appl Stat.* 2003; 52:63–76.
29. Plummer, M. JAGS version 4.2.0 user manual. Lyon, France: International Agency for Research on Cancer; 2016.
30. Plummer, M., Stukalov, A., Denwood, M. *rjags: Bayesian Graphical Models using MCMC.* 2016. Available at: <https://cran.r-project.org/web/packages/rjags/index.html>. Accessed April 3, 2017
31. R Core Team. *R: a language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing; p. 2017 Available at: <http://www.r-project.org>
32. Nunes MS with contributions from T. Heuer, C., Marshall, J., Sanchez, J., Thornton, R., Reiczigel, J., et al. *epiR: Tools for the Analysis of Epidemiological Data.* 2016. Available at: <https://cran.r-project.org/web/packages/epiR/index.html>. Accessed April 3, 2017
33. Colby, SL., Ortman, JM. Projections of the Size and Composition of the US Population: 2014 to 2060. 2015. Available at: <https://www.census.gov/library/publications/2015/demo/p25-1143.html>. Accessed April 6, 2017
34. Jafarzadeh SR, Felson DT. Corrected estimates for the prevalence of self-reported doctor-diagnosed arthritis among US adults. *Arthritis Rheumatol* Hoboken NJ. 2017
35. Badley EM, Canizares M, Perruccio AV. A population-based study of changes in arthritis prevalence and arthritis risk factors over time: Generational differences and the role of obesity. *Arthritis Care Res.* 2017
36. Bolen J, Helmick CG, Sacks JJ, Gizlice Z, Potter C. Should people who have joint symptoms, but no diagnosis of arthritis from a doctor, be included in surveillance efforts? *Arthritis Care Res.* 2011; 63:150–154.
37. Feinglass J, Nelson C, Lawther T, Chang RW. Chronic joint symptoms and prior arthritis diagnosis in community surveys: implications for arthritis prevalence estimates. *Public Health Rep Wash DC* 1974. 2003; 118:230–239.
38. Busija L, Buchbinder R, Osborne RH. Quantifying the impact of transient joint symptoms, chronic joint symptoms, and arthritis: a population-based approach. *Arthritis Rheum.* 2009; 61:1312–1321. [PubMed: 19790129]
39. 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]
40. Neyman J. On the two different aspects of the representative method: The method of stratified sampling and the method of purposive selection. *J R Stat Soc.* 1934; 97:558–625.

41. Gelman A. Struggles with Survey Weighting and Regression Modeling. *Stat Sci.* 2007; 22:153–164.
42. Rao JNK. Impact of Frequentist and Bayesian Methods on Survey Sampling Practice: A Selective Appraisal. *Stat Sci.* 2011; 26:240–256.
43. Little RJ. Calibrated Bayes, an alternative inferential paradigm for official statistics. *J Off Stat.* 2012; 28:309.
44. Bell RM, Cohen ML. Comment: Struggles with Survey Weighting and Regression Modeling. *Stat Sci.* 2007; 22:165–167.
45. Yelin E, Murphy L, Cisternas MG, Foreman AJ, Pasta DJ, Helmick CG. Medical care expenditures and earnings losses among persons with arthritis and other rheumatic conditions in 2003, and comparisons with 1997. *Arthritis Rheum.* 2007; 56:1397–1407. [PubMed: 17469096]

Cross-classified outcomes of arthritis surveillance criteria based on doctor-diagnosed arthritis, chronic joint symptoms, and symptoms onset.

TABLE 1

Sub-Population	DDx	CJS	S-3M	Frequency (%)	DDx+ (%)	CJS+, DDx- (%)	Total (%)
N = 33,672							
Sex							
Age (Year)							
Men	18-64			1,740 (15.0)	2,242 (19.3)	11,597 (34.4)	
		+	+	1,260 (3.7)			
		+	.	405 (1.2)			
		+	-	75 (0.2)			
		-	+	1,849 (5.5)			
		-	.	7,615 (22.6)			
		-	-	393 (1.2)			
Women	18-64			2,734 (20.0)	2,294 (16.7)	13,697 (40.7)	
		+	+	2,002 (5.9)			
		+	.	608 (1.8)			
		+	-	124 (0.4)			
		-	+	1,856 (5.5)			
		-	.	8,669 (25.7)			
		-	-	438 (1.3)			
Men	65			1,511 (43.5)	545 (15.7)	3,474 (10.3)	
		+	+	980 (2.9)			
		+	.	469 (1.4)			
		+	-	62 (0.2)			
		-	+	477 (1.4)			
		-	.	1,418 (4.2)			
		-	-	68 (0.2)			
Women	65			2,704 (55.1)	660 (13.5)	4,904 (14.6)	
		+	+	1,958 (5.8)			
		+	.	608 (1.8)			
		+	-	138 (0.4)			
		-	+	581 (1.7)			

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Sub-Population	DDx	CJS	S-3M	Frequency (%)	DDx+ (%)	CJS+, DDx- (%)	Total (%)
Sex	-	-	.	1,540 (4.6)			N = 33,672
Age (Year)	-	+	-	79 (0.2)			

The null value (.) indicates that the value for the symptoms onset was not available due to a negative affirmation for recent chronic joint symptoms criterion.
 CJS: chronic joint symptoms; DDx: doctor-diagnosed arthritis; S-3: symptoms onset exceeded 3 months

TABLE 2

Posterior probability estimates for the measurement error-adjusted prevalence of arthritis among US adults in 2015.

Sex	Age (Year)	Posterior Median (95% PI)	
		Model with Distinct Sensitivity for 4 Sub-Populations Stratified by Age and Sex	Model with Identical Sensitivity for Men and Women, and Distinct Sensitivity by Age
Men	18–64	29.9% (23.4%, 42.3%)	24.3% (18.3%, 32.3%)
Women	18–64	31.2% (25.8%, 44.1%)	34.0% (25.8%, 44.5%)
Men	65	55.8% (49.9%, 70.4%)	57.9% (50.6%, 65.3%)
Women	65	68.7% (62.1%, 79.9%)	75.8% (66.6%, 84.6%)

PI: probability interval.

Adjusted prevalence, in contrast to unadjusted prevalence, is the estimate that is corrected for the measurement errors as a result of imperfect sensitivity and specificity of the surveillance criteria.

TABLE 3

Accuracy measures for the arthritis surveillance criteria.

Criterion	Parameter	Sex	Age (Year)	Posterior Median (95% PI)	
				Model with Distinct Sensitivity for 4 Sub-Populations Stratified by Age and Sex	Model with Identical Sensitivity for Men and Women, and Distinct Sensitivity by Age
DDx	Se	Men	18-64	22.0% (11.1%, 48.3%)	50.0% (39.0%, 64.2%)
	Se	Women	18-64	34.1% (24.3%, 62.3%)	Identical to above
	Se	Men	65	67.9% (59.4%, 76.1%)	71.5% (64.5%, 80.5%)
	Se	Women	65	74.9% (67.4%, 82.0%)	Identical to above
CIS	Sp	Both	Both	87.2% (82.5%, 96.1%)	95.8% (94.2%, 97.2%)
	Se	Men	18-64	62.7% (43.6%, 72.5%)	46.8% (38.2%, 60.3%)
	Se	Women	18-64	64.9% (46.1%, 73.9%)	Identical to above
	Se	Men	65	69.1% (55.5%, 75.4%)	64.2% (59.7%, 70.2%)
S-3M	Se	Women	65	74.1% (64.9%, 80.5%)	Identical to above
	Sp	Both	Both	83.3% (75.1%, 91.8%)	74.8% (72.1%, 77.3%)
	Se	Men	18-64	87.4% (79.5%, 94.4%)	94.1% (93.3%, 94.9%)
	Se	Women	18-64	88.8% (80.4%, 94.7%)	Identical to above
	Se	Men	65	93.1% (89.1%, 94.9%)	94.3% (93.6%, 95.1%)
	Se	Women	65	93.4% (90.8%, 95.1%)	Identical to above
	Sp	Both	Both	15.2% (1.6%, 51.3%)	18.8% (17.5%, 20.7%)

CIS: chronic joint symptoms, DDx: doctor-diagnosed arthritis; PI: probability interval; S-3M: symptoms onset exceeded 3 months; Se: sensitivity; Sp: specificity.