Validation of a Surveillance Case Definition for Arthritis

JEFFREY J. SACKS, LESLIE R. HARROLD, CHARLES G. HELMICK, JERRY H. GURWITZ, SRINIVAS EMANI, and ROBERT A. YOOD

ABSTRACT. ***Objective.*** To assess whether self-reports of chronic joint symptoms or doctor-diagnosed arthritis can validly identify persons with clinically verifiable arthritis.

Methods. The Behavioral Risk Factor Surveillance System (BRFSS), a telephone health survey, defines a case of arthritis as a self-report of chronic joint symptoms (CJS) and/or doctor-diagnosed arthritis (DDx). A sample of health plan enrollees aged 45–64 years and ≥ 65 years with upcoming annual physical examinations were surveyed by telephone using the 2002 BRFSS CJS and DDx questions. Based on responses (CJS+, DDx–; CJS–, DDx+; CJS+, DDx+; CJS–, DDx–), respondents were recruited to undergo a standardized clinical history and physical examination for arthritis (the gold standard for clinical validation). Weighted sensitivities and specificities of the case definition were calculated to adjust for sampling.

Results. Of 2180 persons completing the telephone questionnaire, 389 were examined; of these, 258 met the case definition and 131 did not. For those examined and aged 45 to 64 years (n = 179), 96 persons had arthritis confirmed, of whom 76 met the case definition. Among those examined and aged ≥ 65 (n = 210), 150 had arthritis confirmed, of whom 124 met the case definition. Among those without clinical arthritis, 45 of 83 of those aged 45 to 64 years and 40 of 60 of those aged ≥ 65 did not meet the case definition. For those aged 45 to 64 years, the weighted sensitivity of the case definition in this sample was 77.4% and the weighted specificity was 58.8%; for those aged ≥ 65, the sensitivity was 83.6% and specificity 70.6%. CJS+ had higher sensitivity and lower specificity than DDx+ in the younger age group; CJS+ and DDx+ behaved more comparably in the older age group.

Conclusion. The case definition based on self-reported CJS and/or DDx appeared to be sensitive in identifying arthritis, but specificity was lower than desirable for those under age 65 years. Better methods of ascertaining arthritis by self-report are needed. Until then, a change in the surveillance case definition for arthritis appears warranted. (J Rheumatol 2005;32:340–7)

Key Indexing Terms:

ARTHRITIS	SURVEILLANCE	PREVALENCE
CASE DEFINITIONS	EPIDEMIOLOGY	VALIDATION

Arthritis and other rheumatic conditions, here referred to as arthritis, make up the most frequent cause of disability in the United States[1]. Fortunately, arthritis can be successfully managed, and its effects lessened. Exercise, weight loss, medications, surgery, and educational/sociobehavioral interventions can decrease pain and improve physical function and quality of life[2-6]. A recent report from the Centers for Disease Control and Prevention (CDC) suggests that the number of persons affected by arthritis and aged 65 and older will nearly double by 2030 if arthritis prevalence rates remain stable, owing to the aging of the population[7]. The increasing prevalence of arthritis, the enormous disability associated with these conditions, the substantial healthcare costs (total costs exceeding $86 billion in 1997)[8], and the underutilization of effective interventions[9-11] together make a strong case for public health surveillance of the prevalence and impact of the condition to help direct policy and resources and to monitor achievement of national and state health objectives regarding arthritis[12].

Estimating the prevalence of arthritis in the population can be difficult. Using an International Classification of Disease-based case definition developed by the National Arthritis Data Workgroup[13], Rao and colleagues found that 16.4% of persons who reported having arthritis never sought attention from a physician for the problem, even though

From the Division of Adult and Community Health (DACH), National Center for Chronic Disease Prevention and Health Promotion (NCCDPHP), Centers for Disease Control and Prevention (CDC), Atlanta, Georgia; Meyers Primary Care Institute, Fallon Foundation, and University of Massachusetts Medical School, Worcester, Massachusetts; Fallon Clinic Research Department, West Boylston, Massachusetts; and Fallon Clinic, Worcester, Massachusetts, USA.

J.J. Sacks, MD, MPH; C.G. Helmick, MD, Division of Adult and Community Health, National Center for Chronic Disease Prevention and Health Promotion, Centers for Disease Control and Prevention; L.R. Harrold, MD, MPH; J.H. Gurwitz, MD, Meyers Primary Care Institute; S. Emani, PhD, Fallon Clinic Research Department; R.A. Yood, MD, Fallon Clinic.

Address reprint requests to Dr. J.J. Sacks, Centers for Disease Control and Prevention, 4770 Buford Highway, MS-K51, Atlanta, GA 30341-3724. E-mail: jjs3@cdc.gov

Submitted April 19, 2004; revision accepted September 21, 2004.

Personal non-commercial use only. The Journal of Rheumatology Copyright © 2005. All rights reserved.

Downloaded on April 19, 2024 from www.jrheum.org

nearly three-quarters of these persons had one or more doctor visits within the preceding 12 months[14]. Rao, *et al* concluded that nearly 6 million Americans with self-reported arthritis never see a physician for their condition.

Because medical encounter and claims-based data cannot record data about persons who have never seen a physician for their arthritis symptoms, self-report surveys have been used to ascertain a more complete estimate of the burden of arthritis on the American public. In particular, the state-based telephone survey — the Behavioral Risk Factor Surveillance System (BRFSS) — has used questions on chronic joint symptoms (CJS) and whether persons had a previous diagnosis of arthritis by a doctor (DDx) to define a case of arthritis for the purpose of public health surveillance.

The questions on CJS were originally developed in the mid-1990s by the National Arthritis Data Workgroup. The question on DDx was developed by state BRFSS coordinators in the mid 1990s. Because of growing concerns with estimate stability, cognitive performance, and evidence that acute, self-limited injuries were being classified as cases of arthritis, both sets of questions were revised in 2001, cognitively tested, and used for the first time in 2002.

Validation of the surveillance case definition based on self-reported information about CJS and DDx has not been done and is essential if this definition is going to be used for disease surveillance and to analyze temporal trends at national and state levels, as well as for epidemiologic research. This study uses the results from a standardized clinical history and physical examination to define clinically significant arthritis as the gold standard to validate the case definition for arthritis based on self-reported CJS or DDx from a telephone survey using the revised version of the arthritis questions.

MATERIALS AND METHODS

The study was conducted at the Fallon Clinic, Inc., a multispecialty group practice that operates 27 medical centers throughout central Massachusetts. Fallon Clinic medical staff provide care for about 130,000 members of the Fallon Community Health Plan (FCHP), a nonprofit health maintenance organization. The FCHP has a computerized database that includes demographic information, outpatient encounters and diagnoses, pharmacy use, and hospitalizations.

The respondent set was FCHP members aged 45 years and older who had scheduled annual physical examinations with their primary care physician at any one of 3 Fallon Clinic sites. After identification through a computerized appointment system, a letter was sent to these patients 4 weeks before their scheduled visit informing them of the study and that they would be called for a 10 minute telephone interview in the next few days. The letter advised patients that they could call and decline to take part in the study.

Interviewers made 3 to 5 attempts to call patients. The times and days were varied and attempts were also made on weekends. The telephone survey contained the revised questions on CJS and DDx used in the 2002 BRFSS survey. Because the intent was to administer the questions in the same manner as they are typically given in the field conditions of BRFSS use, no special attention was drawn to the arthritis questions; they appeared with additional questions on a number of other content areas, such as hypertension, asthma, diabetes, tobacco use, and exercise, so as not to overly focus on arthritis. In addition, there were questions assessing functional status from the Multidimensional Health Assessment Questionnaire (MDHAQ)[15].

CJS+ was defined as the respondent answering yes to the following 2 questions: (1) "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?" and (2) "Did your joint symptoms first begin more than 3 months ago?" DDx+ was defined as the respondent answering yes to, "Have you ever been told by a doctor or other health professional that you have some form of arthritis, rheumatoid arthritis (RA), gout, lupus, or fibromyalgia?" The case definition for arthritis was considered as CJS+ and/or DDx+.

The selection of persons for the clinical examination for arthritis was based on age group (45–64 years and ≥ 65 years) and responses to questions on CJS and DDx. We aimed to perform a standardized clinical history and physical examination on ~50 persons in each of 4 self-reported response categories: chronic joint symptoms without doctor-diagnosed arthritis (CJS+DDx–), doctor-diagnosed arthritis without chronic joint symptoms (CJS–DDx+), both chronic joint symptoms and doctor-diagnosed arthritis (CJS+DDx+), and neither chronic joint symptoms nor doctor-diagnosed arthritis (CJS–DDx–), for a total of 400 patients to be examined. A target of 50 for each cell was chosen because (1) we wanted to be certain to test the performance of each combination of age and response; and (2) sample sizes of roughly 40 usually produce normal approximations for binomial distributions. At the end of the telephone survey, depending on the combinations of responses and the recruitment needs at the time, we varied the closing script to invite some of the respondents for the special examination and not others. Selected respondents were told this undertaking was a study of arthritis and were invited to have an additional 30 minute history and physical examination at their upcoming scheduled appointment.

At the time of the scheduled appointment, written consent was obtained and one of 2 specially trained nurses provided patients with a printed questionnaire. This included questions on CJS and DDx (to check reproducibility of answers), the Health Assessment Questionnaire (HAQ)[16], advanced activities of daily living scale from the MDHAQ, average level of pain, prior diagnoses, limitation, current treatment, and other clinical details. Patients who completed the clinical evaluation received a $20 gift certificate.

After the patient completed the written questionnaire, the research nurse conducted a standardized clinical history and physical examination. The nurse also reviewed medical records, including physician notes, laboratory results, and radiographic procedures, to help decide if the person had clinically significant arthritis or other rheumatic condition. Nurses were not aware of which response category the study subject was from, i.e., whether the patient was CJS or DDx positive or negative from the telephone survey. A rheumatologist then reviewed the nurse's written evaluation of each patient and nurse-selected relevant medical records to verify the nurse's diagnosis. The rheumatologist's determination was considered the gold standard to confirm the BRFSS case definition.

Conditions considered to represent and confirm the case definition were those designated by the National Arthritis Data Workgroup[13]. To be considered a gold-standard case of arthritis, a condition had to have "clinical significance," i.e., it had to be either symptomatic or something a physician would treat. Thus, a person with asymptomatic Heberden's nodes who required no treatment was not considered to have clinically significant arthritis. Similarly, a patient with a joint radiograph obtained for evaluation of an injury would not be considered as having clinically significant arthritis if symptoms resolved after the injury even if the radiograph had shown underlying osteoarthritis (OA). Someone with symptomatic arthritis or bursitis, even if not undergoing treatment, was considered to have clinically significant arthritis. Persons with neck or back pain were not considered to have clinically significant arthritis unless there was radiographic documentation of spinal arthritis in the medical records and it was felt the etiology of the pain was from the arthritis. Asymptomatic patients with episodic arthritis in the past and at risk for recurrent symptoms were considered to

Personal non-commercial use only. The Journal of Rheumatology Copyright © 2005. All rights reserved.

Downloaded on April 19, 2024 from www.jrheum.org

have clinically significant arthritis. For example, patients with prior gout attacks were considered to have clinically significant arthritis because of their risk for future attacks of gout. A patient whose knee pain had previously been attributed to OA documented by radiograph was considered to have clinically significant arthritis even if asymptomatic at the time of the nurse evaluation.

Prior to the study, a board certified rheumatologist (RY) trained both nurses in the history and examination skills necessary to diagnose arthritis or other rheumatic conditions. This training included reading a training manual on evaluating potential arthritis patients, a videotape on performing a musculoskeletal examination, and spending 16 hours in the rheumatology clinic. The first 8 hours involved observing the rheumatologist evaluating patients, and the second 8 hours involved the nurses evaluating patients under the rheumatologist's supervision. In addition to the rheumatologist reviewing all written results of all assessments by the nurses, a special interrater reliability study was performed.

Interrater reliability study. A 10% sample of each nurse's patients underwent a separate complete examination by one of 3 rheumatologists who had no knowledge of the nurse's assessment. The rheumatologists saw the subject directly after the nurse and determined whether the subject had or did not have clinically significant arthritis. Each rheumatologist saw between 13 and 16 such patients and filled out the same study form used by the nurse for the examination. Kappa statistics were calculated for each nurse[17] to determine agreement between the nurse's and the rheumatologist's evaluations.

Calculation of weighted sensitivity and specificity. To determine whether the case definitions based on responses to CJS and/or DDx in the telephone survey accurately identified patients with clinically significant arthritis, the age- and CJS/DDx response-specific cell sampling fractions were determined and used to weight those cells back to the Fallon Clinic population. Sensitivity, specificity, and positive predictive value of the case definition, self-reported CJS, and self-reported DDx were calculated within age groups.

Study protocols were approved by the Saint Vincent Hospital-Fallon Clinic-Fallon Community Health Plan Institutional Review Board and the CDC Human Subjects Review Board. Telephone calls and clinical examinations occurred between November 2001 and October 2002.

RESULTS

Response rates. Attempts were made to telephone 1816 persons aged 45–64 years scheduled for an annual physical examination. Of those called, 437 refused; 302 were not available for interview before their annual physical; 43 telephone numbers were wrong, businesses, no answer, answering machine, or nonworking; 4 had incomplete nonusable interviews; and 10 had language barriers, leaving 1020 completed telephone interviews (response rate = 57.3%).

Attempts were made to telephone 2015 persons 65 years and older scheduled for an annual examination. Of those called, 619 refused; 150 were not available for interview before their annual physical; 46 telephone numbers were wrong, businesses, no answer, answering machine, or nonworking; 2 had incomplete nonusable interviews; 37 had language barriers; and one was deceased, leaving 1160 completed telephone interviews (response rate = 58.5%).

It was not possible to calculate the proportion of people offered an examination who agreed or refused to take it. Once the desired cell size of 50 was reached for a particular age and specific case definition cell, persons in that cell identified during the telephone interview were no longer invited to participate in the examination, although because of patients already in the queue we did exceed the 50-person target in some cells. Because of their infrequency in the population, we found it very difficult to identify and recruit persons with a doctor diagnosis but no symptoms (CJS–, DDx+). Eventually, resource, time, and logistic considerations forced us to abandon further recruitment. Based on the cell-specific numbers who completed a telephone interview and then were examined, it is clear that cell-specific examination rates were at least 14%–28% (Table 1), and that unequal weights needed to be applied to represent the population.

Demographic comparability of those examined with those not examined. No significant differences were found between those interviewed by telephone and not examined compared with those interviewed by telephone and examined, except for a lower percentage reporting nonoccupational physical activity (NOPA) in the past 30 days (72.7% vs 77.7%; $p < 0.05$; Table 2).

Validating the case definition. Overall, about 2 in 3 of those 2180 interviewed by telephone met the case definition, i.e., 614 (60.7%) of those interviewed 45 to 64 years old and 795 (69.3%) of those aged ≥ 65 years. Of those interviewed by telephone, 389 were examined; of these, 258 met the case definition and 131 did not. For those examined aged 45 to 64 years (n = 179), 96 persons had clinically significant arthritis, of whom 76 met the case definition. Among those examined aged ≥ 65 years (n = 210), 150 had clinically significant arthritis, of whom 124 met the case definition. Among those without clinically significant arthritis, 45 of 83 of those aged 45 to 64 years and 40 of 60 of those aged ≥ 65 years did not meet the case definition. The highest unweighted rates of clinical confirmation occurred among the CJS+DDx+ group, of whom 81% aged 45–64 years and 100% of those ≥ 65 years had clinically significant arthritis (Table 3). Although the unweighted data in Table 3 suggest that rates of not being clinically confirmed (false positive) are roughly equal in the CJS+DDx– and CJS–DDx+ groups among those 45–64 years of age (42% and 46%, respectively), it is important to appreciate that the CJS–DDx+ stratum in this age group is the least frequent in the whole population (Table 1), and thus contributes relatively less to the weighted estimates of sensitivity and specificity for the overall case definition or components of the definition (shown below).

False positives (case definition positive, clinical examination negative). The 58 false positives were very unlikely to be CJS+DDx+ (19% for ages 45–64 years and 0% for ≥ 65 years), but were most likely when only one component of the definition, i.e., CJS+ alone or DDx+ alone, was reported by the patient (Table 3). About 90% of the persons with false positives reported their joint symptoms began more than one year prior to telephone interview. The median time of onset of CJS was 2 years for true positives and 3 years for false

Personal non-commercial use only. The Journal of Rheumatology Copyright © 2005. All rights reserved.

Downloaded on April 19, 2024 from www.jrheum.org

Table 1. Numbers examined by age group, weight factors, and telephone categorization.

	Categorization by Telephone Question		No. of Patients		
	CJS+	DDx+	Completed Interview	Took Examination	Weight Factor
Age Group					
45–64	Yes	No	234	43	5.442
	No	Yes	85	24	3.542
	Yes	Yes	297	47	6.319
	No	No	400	65	6.154
≥ 65	Yes	No	158	44	3.591
	No	Yes	178	34	5.235
	Yes	Yes	467	66	7.076
	No	No	356	66	5.394

Table 2. Demographic characteristics of 2175* interviewed study subjects, by examination status.

	Interviewed But Not Examined (n = 1786)* No. (%)	Interviewed and Examined (n = 389) No. (%)
Age		
45–64 yrs	837 (46.9)	179 (46.0)
≥ 65 yrs	949 (53.1)	210 (54.0)
Female	1001 (56.0)	211 (54.2)
White	1744 (98.4)	379 (97.4)
Non-Hispanic	1743 (99.4)	377 (99.2)
Education		
Some high school or less	353 (19.8)	74 (19.8)
High school or graduate equivalency diploma	733 (41.1)	148 (39.6)
Some college or more	695 (39.0)	152 (40.6)
Marital status		
Married	1213 (68.1)	271 (72.5)
Widowed	322 (18.1)	57 (15.2)
Divorced, never married, or other	246 (13.8)	46 (12.3)
Employment		
Retired	904 (50.7)	195 (52.0)
Employed or self-employed	779 (43.7)	156 (41.6)
Out of work, homemaker, student, or unable to work	103 (5.7)	24 (6.4)
Health status		
Excellent	242 (13.6)	63 (16.2)
Very good	536 (30.1)	115 (29.6)
Good	656 (36.8)	153 (39.4)
Fair	294 (16.5)	46 (11.9)
Poor	54 (3.0)	11 (2.8)
Participated in nonoccupational physical activity in past 30 days	1257 (72.2)	300 (77.7)**
Body Mass Index class		
Normal	615 (35.7)	133 (35.7)
Overweight	695 (40.4)	160 (42.9)
Obese	411 (23.9)	80 (21.5)
Meets case definition	1161 (65.0)	258 (66.3)
CJS+	956 (53.5)	200 (51.4)
DDx+	856 (47.9)	171 (44.0)

* The results of 5 completed interviews were lost. ** Statistically significant difference ($p < 0.05$).

positives; time of onset did not differ significantly by age group. The median joint pain and aching level in the past month reported on a 10-point scale was higher for younger and older true positives (4, 4) and lower for younger false positives and older false positives (2, 1, respectively). There were a number of additional differences between true posi-

Personal non-commercial use only. The Journal of Rheumatology Copyright © 2005. All rights reserved.

Downloaded on April 19, 2024 from www.jrheum.org

Table 3. Results of gold-standard clinical examination, by self-reported chronic joint symptoms (CJS) and doctor diagnosis (DDx), by age group.

	Telephone Response Category	Clinically Significant Arthritis by Standardized History and Physical Examination: Yes No. (%)	No No. (%)
Age Group			
45–64 yrs	Case (CJS+DDx–)	25 (58)	18 (42)
	Case (CJS–DDx+)	13 (54)	11 (46)
	Case (CJS+DDx+)	38 (81)	9 (19)
	Not case (CJS–DDx–)	20 (31)	45 (69)
≥ 65 yrs	Case (CJS+DDx–)	35 (80)	9 (20)
	Case (CJS–DDx+)	23 (68)	11 (32)
	Case (CJS+DDx+)	66 (100)	0 (0)
	Not case (CJS–DDx–)	26 (39)	40 (61)

Data are unweighted. Weighted clinical confirmation rates for row combinations by age are:
45–64 years (CJS+DDx+ and CJS+DDx– = 71%, CJS–DDx+ and CJS+DDx+ = 75%, CJS+DDx+, CJS+DDx–, and CJS–DDx+ = 69%);
≥ 65 years (CJS+DDx+ and CJS+DDx– = 95%, CJS–DDx+ and CJS+DDx+ = 91%, CJS+DDx+, CJS+DDx–, and CJS–DDx+ = 89%)

tives and false positives, as indicated in Table 4; true positives were significantly more likely to report more limitations, being seen by a doctor for joint symptoms, and being under treatment for joint symptoms. For both age groups, false positives had the same median age as true positives and the same education level distribution. False positives tended to have better reports of functioning than true positives for all HAQ items, although differences were not statistically significant.

Of 31 persons with false positives who reported a prior doctor diagnosis of arthritis (20 in the younger age group and 11 in the older age group), some provided specific information on the nature of their prior diagnosis — among those aged 45–64, 5 reported OA, 3 RA, and one gout; for the older age group false positives, the reports were one OA and one RA.

False negatives (case definition negative, clinical examina-

Table 4. Comparison of false positives and true positives, by age group.

	Percentage of: Age 45–64 yrs True Positive, n = 76	Age 45–64 yrs False Positive, n = 38	Age ≥ 65 yrs True Positive, n = 124	Age ≥ 65 yrs False Positive, n = 20
Self-Reported Characteristic				
Overall health status reported as fair or poor	18.4	5.3	22.0	10.0
Limited by arthritis or joint symptoms (AJS)				
First report on phone survey	40.3	15.2+	31.4	23.5
Second report on phone survey	48.4	10.0+++	30.8	33.3
Report on written survey	46.1	7.9+++	30.8	10.5
Saw physician for joint symptoms	77.6	53.6+	71.2	33.3+
AJS affects work for pay	24.2	15.6	12.0	0.0
Currently treated by a physician for AJS				
Report on phone survey	42.0	0.0+++	27.5	0.0
Report on written survey	30.0	0.0+++	28.1	0.0++
Doctor recommended taking medications for AJS	60.9	13.2+++	53.8	11.5+++
Took medications for AJS in past 30 days	65.3	21.6+++	61.0	27.8+
Body mass index category of overweight or obese	68.1	62.9	72.1	50.0+++
Female	57.9	57.9	56.5	50.0
Participated in physical activity or exercise past 30 days	69.7	84.2	74.8	75.0

Unknowns excluded from percentage calculations. False positive vs true positive: + $p < 0.05$, ++ $p < 0.01$, +++ $p < 0.001$.

Personal non-commercial use only. The Journal of Rheumatology Copyright © 2005. All rights reserved.

Downloaded on April 19, 2024 from www.jrheum.org

tion positive). The 20 persons with false negatives in the 45–64 age group were more likely than persons with true negatives to report the following on the written questionnaire given at the time of clinical examination: DDx (20% vs 0%; $p < 0.01$) and CJS (50% vs 16%; $p < 0.01$). They were also more likely to be female (75% vs 42%; $p < 0.05$). For the 26 false negatives in the ≥ 65 age group, the only significant difference between them and true negatives was CJS (38% vs 8%; $p < 0.01$) reported at the time of the written questionnaire. For both age groups, false negatives showed very little difference from true negatives for reports of levels of functioning; they were essentially the same for all HAQ items.

Sensitivity, specificity, and positive predictive value of the case definition. For those aged 45–64 years, the weighted sensitivity of the case definition in this sample was 77.4% and the weighted specificity was 58.8%; for those aged ≥ 65, the sensitivity was 83.6% and specificity 70.6% (Table 5). CJS+ had higher sensitivity and lower specificity than DDx+ in the younger age group; CJS+ and DDx+ behaved more comparably in the older age group. Positive predictive value increased with age and was better for the CJS+ or DDx+ components individually rather than the overall case definition formed from the combination of the components.

Reliability of self-reports. For 376 examined persons who completed a written questionnaire at the time of the examination, 135/176 (76.7%) aged 45–64 and 136/200 (68.0%) aged ≥ 65 had exactly matching responses for both CJS and DDx on the telephone and written surveys. Differences in consistent reporting were worse for CJS than for DDx. For those 45–64 years, 11 telephone survey cases were not classified as such on the written survey and 18 non-cases would have been reclassified as cases on the written survey. For those ≥ 65 years, 14 telephone survey cases were not classified as such on the written survey and 16 non-cases would have been reclassified as cases on the written survey.

Interrater reliability. A total of 44 subjects were evaluated by both a nurse (22 with nurse 1 and 22 with nurse 2) and one of 3 rheumatologists. Patients were drawn from the following cells: 11 CJS+DDx–, 10 CJS–DDx+, 12 CJS+DDx+, and 11 CJS–DDx–. In 25 instances the rheumatologist and nurses agreed the subject had clinically significant arthritis; in 16 instances they agreed the subject did not have clinically significant arthritis. In 3 instances the nurses felt the subject had clinically significant arthritis, but the rheumatologist did not. The discrepant observations were all in subects aged 45–64 and examined by one rheumatologist; their telephone survey classifications were one CJS+DDx+, one CJS+DDx–, and one CJS–DDx–. The kappa statistics were 0.82 for nurse 1 and 0.90 for nurse 2, indicating very good agreement.

DISCUSSION

Our study addresses an important need identified in previous studies using BRFSS, that of validating the surveillance case definition for arthritis based on self-report[18]. Our findings suggest that although the surveillance case definition based on self-reported CJS and/or DDx appeared sensitive in this population in identifying clinically significant arthritis, the specificity was lower than desirable for those under age 65 years for the purpose of accurately assessing population disease burden. Disaggregation of the components of the case definition suggested CJS and DDx subjects behave differently by age. Used together, the components improve the sensitivity of the case definition over either component alone; however, the sensitivity improvement comes at the cost of lowered specificity.

Specificity was lowered because of persons with false positives, who in this study had similar demographics, but better health, less pain, and fewer limitations than persons with true positives. Specificity was higher for either component of the case definition alone compared to the combination of the components.

Our gold standard for these estimates of sensitivity and specificity is subject to certain limitations. While our study suggests that trained nurses can do an examination for clinically significant arthritis with good accuracy, it is not clear that a standardized clinical history and physical examination is the *sine qua non* for determining if a person has arthritis or not. Clinical examinations are subject to false negatives and false positives. We did not perform any laboratory or radiographic studies of patients to evaluate them for arthritis, but relied on the history, clinical examination, and information in the medical records. Because of the episodic

Table 5. Weighted* sensitivity, specificity, and positive predictive value (PPV) of the case definition, by age and component of the case definition.

	Age Group, yrs	Sensitivity	Specificity	PPV
Case definition	45–64	77.4	58.8	68.5
	≥ 65	83.6	70.6	88.9
CJS+	45–64	69.0	67.1	70.8
	≥ 65	69.5	89.4	94.8
DDx+	45–64	52.5	79.6	74.9
	≥ 65	68.8	81.1	91.0

* Data weighted to the Fallon Clinic population.

Personal non-commercial use only. The Journal of Rheumatology Copyright © 2005. All rights reserved.

Downloaded on April 19, 2024 from www.jrheum.org

process of some arthritic conditions, patients with no symptoms or clinical findings at the time of their examination might have had arthritis. A patient complaining of symptoms in the absence of signs may also have arthritis, especially in early stages. While persons with false positives in this study had longer duration of symptoms than persons with true positives (median one year earlier onset), their average pain level was far less than true positives. It is possible that some of those classified as false positive had mild disease that was not apparent on clinical examination and thus were misclassified on the gold standard examination. Indeed, if the gold standard is too restrictive and excludes true cases of arthritis correctly identified by the surveillance definition of arthritis, then true cases would be classified as false positives and produce a falsely low estimate of specificity.

Our results are subject to a number of other limitations, including many sources of variation in the results. First, the study population may not be highly generalizable — it was relatively small and homogeneous, i.e., primarily non-Hispanic whites; results in other population groups may differ. The examined sample did appear representative of the telephoned sample; the only difference (nonoccupational physical activity), while statistically significant, was not a clinically significant difference in our view. Second, we did not study persons under 45 years of age and the data suggest there are age effects on sensitivity and specificity of the case definition. Third, it is clear from our data that there were changes in study subject responses on CJS and DDx (and thus, their classification as a surveillance case or not) between the time of the telephone survey and written questionnaire at the time of the examination (about 3 weeks apart). We found a discordance rate of 19% between responses to the CJS questions and 12% for responses to the doctor diagnosis question on the telephone and written surveys. While it could be argued that because CJS is based on 2 questions rather than one, and because symptoms could change in the interim, CJS responses might vary between the telephone and the written survey, the discordance in response to the doctor diagnosis question is remarkable. Of 382 examinees providing responses to the same doctor diagnosis question on the telephone and written surveys, the discordant responses were as follows: 23 were DDx– telephone/DDx+ written, 17 were DDx+ telephone/DDx– written, and 7 reported "don't know" on the telephone and DDx positive or negative on the written questionnaire. The extent to which understanding and comprehension related to the modes of administration (telephone vs written) account for such differences in response is unclear. The discordance rate was not related to age.

Nevertheless, this was a rigorous study of almost 400 individuals. We found that the case definition detected many more cases than were found to have clinically significant arthritis by history and physical examination, i.e., we found many false positives from the case definition — a specificity-related issue. Sensitivity and specificity typically interact inversely; when one increases, the other decreases. Thus, there is a tradeoff to be made between detecting cases (sensitivity) and properly identifying those without the condition of interest (specificity). When the failure to detect a "case" is costly, e.g., missing a case of measles can spawn a chain of further transmission, high sensitivity is valued over high specificity in public health surveillance. When surveillance is primarily used to monitor disease burden and allocate resources in a population, higher specificity may be preferable to higher sensitivity, even with a common condition such as arthritis[19].

Because the pool of adults 18 years of age and older in the US population (the target of the BRFSS) who do not truly have arthritis is so much larger than the number who do, for all equal values of sensitivity and specificity, the number of false positives produced would be much larger than the number of false negatives. Using the higher sensitivity and lower specificity values from Table 5 exacerbates the difference, especially in the population aged 45–64 years, and likely by extension in those 18–44 years old. Including these many false positives may inflate numbers of estimated cases unreasonably, strain credibility, and misdirect arthritis programs and resources to populations not in need of interventions. Accordingly, we conclude that better methods of ascertaining arthritis by self-report are needed. Until those methods are developed, a change in the case definition for arthritis surveillance appears warranted.

ACKNOWLEDGMENT

The authors thank Candace LeBlanc, RN, and Marcia Kirkpatrick, RN, for their expert nurse evaluations of the study subjects, Nancy Walker, MD, MPH, for participating in the interrater reliability study and review of nurse evaluation forms, Ellen Trencher for organizational support, and Gary Langmaid for computer programming.

REFERENCES

1. Centers for Disease Control and Prevention. Prevalence of disabilities and associated health conditions among adults — United States, 1999. Morb Mortal Weekly Rep 2001;50:120-5.
2. Hochberg MC, Altman RD, Brandt KD, et al. Guidelines for medical management of osteoarthritis. Arthritis Rheum 1995;38:1535-46.
3. Minor MA, Hewett JE, Webel R, Anderson SK, Kay DR. Efficacy of physical conditioning exercise in rheumatoid arthritis and osteoarthritis. Arthritis Rheum 1989;32:1396-405.
4. Felson DT, Zhang Y, Anthony JM, Naimark A, Anderson JJ. Weight loss reduces the risk for symptomatic knee osteoarthritis in women. The Framingham Study. Ann Intern Med 1992;116:535-9.
5. Egsmose C, Birger L, Borg G, et al. Patients with rheumatoid arthritis benefit from early 2nd line therapy: 5 year followup of a prospective double blind placebo controlled study. J Rheumatol 1995;22:2208-13.
6. Lorig KR, Mazonson PD, Holman HR. Evidence suggesting that health education for self-management in patients with chronic arthritis has sustained health benefits while reducing health care

Personal non-commercial use only. The Journal of Rheumatology Copyright © 2005. All rights reserved.

Downloaded on April 19, 2024 from www.jrheum.org

costs. Arthritis Rheum 1993;36:439-46.
7. Centers for Disease Control and Prevention. Projected prevalence of self-reported arthritis or chronic joint symptoms among persons aged 65 years — United States, 2005-2030. Morb Mortal Weekly Rep 2003;52:489-91.
8. Yelin E, Cisternas MG, Pasta DJ, Trupin L, Murphy L, Helmick CG. Medical care expenditures and earning losses of persons with arthritis and other rheumatic conditions in 1997: total and incremental estimates. Arthritis Rheum 2004;50:2317-26.
9. Bellamy N, Bradley LA. Workshop on chronic pain, pain control and patient outcomes in rheumatoid arthritis and osteoarthritis. Arthritis Rheum 1996;39:357-62.
10. Katz JN. Preferences, quality, and the (under)utilization of total joint arthroplasty. Med Care 2001;39:203-5.
11. Hootman JM, Helmick CG, Schappert SM. Magnitude and characteristics of arthritis and other rheumatic conditions on ambulatory medical care visits, United States, 1997. Arthritis Rheum 2002;47:571-81.
12. Meenan RF, Callahan LF, Helmick CG. The National Arthritis Action Plan: A public health strategy for a looming epidemic. Arthritis Care Res 1999;12:79-81.
13. Helmick CG, Lawrence RC, Pollard RA, Lloyd E, Heyse SP. Arthritis and other rheumatic conditions: who is affected now. Who will be affected later. Arthritis Care Res 1995;8:203-11.
14. Rao JK, Callahan LF, Helmick CG. Characteristics of persons with self-reported arthritis and other rheumatic conditions who do not see a doctor. J Rheumatol 1997;24:169-73.
15. Pincus T, Swearington C, Wolfe F. Toward a Multidimensional Health Assessment Questionnaire (MDHAQ). Assessment of advanced activities of daily life and psychological status in the patient-friendly Health Assessment Questionnaire format. Arthritis Rheum 1999;42:2220-30.
16. Ramsey DR, Raynauld J, Fries JF. The Health Assessment Questionnaire 1992: status and review. Arthritis Care Res 1992;5:119-29.
17. Sackett DL, Haynes RB, Guyatt GH, Tugwell P, editors. Clinical epidemiology: a basic science for clinical medicine. 2nd ed. Boston: Little, Brown and Company; 1991:25-31.
18. Mili F, Helmick CG, Zack MM, Moriarty DG. Health related quality of life among adults reporting arthritis: analysis of data from the Behavioral Risk Factor Surveillance System, US, 1996-99. J Rheumatol 2003;30:160-6.
19. LaValley M, McAlindon TE, Evans S, Chaisson CE, Felson DT. Problems in the development and validation of questionnaire-based screening instruments for ascertaining cases with symptomatic knee osteoarthritis. The Framingham Study. Arthritis Rheum 2001;44:1105-13.

Personal non-commercial use only. The Journal of Rheumatology Copyright © 2005. All rights reserved.

Downloaded on April 19, 2024 from www.jrheum.org