

## PRACTICE OF EPIDEMIOLOGY

### Validation of Self-reported Chronic Obstructive Pulmonary Disease in a Cohort Study of Nurses

R. Graham Barr,<sup>1-3</sup> Julie Herbstman,<sup>1</sup> Frank E. Speizer,<sup>1,4</sup> and Carlos A. Camargo, Jr.<sup>1,3,5</sup>

Chronic obstructive pulmonary disease (COPD) is the fourth leading cause of death, and its prevalence is increasing; however, few strategies are available for disease prevention or modification besides smoking interventions. To facilitate examination of modifiable risk factors for COPD in the Nurses' Health Study, the authors validated a questionnaire-based definition of COPD in a subset of this US cohort. Participants were surveyed biennially about lifestyle factors, including smoking, since 1976 and physician diagnosis of COPD since 1988. Self-reported cases were defined as reporting COPD on both the original (1988–1996) and supplemental (1998) questionnaires. The authors requested medical records for a 10% random sample of 2,790 cases and reviewed these records in a systematic, blinded fashion. Validated cases required obstructive spirometry, emphysema on chest radiograph or computed tomography, or physician diagnosis. COPD was confirmed for 78% of 273 cases. Spirometry or radiographic results were available for 84%; when available, mean forced expiratory volume in 1 second was 51% predicted (standard deviation, 19). Applying these results to a hypothetical cohort, the authors estimated the degree to which disease misclassification biases relative risks toward the null value, confirming that questionnaire-based COPD research should focus on minimizing false positives rather than false negatives. In conclusion, COPD can be studied in large, questionnaire-based cohorts of health professionals. *Am J Epidemiol* 2002;155:965–71.

airway obstruction; cohort studies; epidemiologic methods; lung diseases, obstructive; pulmonary disease, chronic obstructive; questionnaires

Chronic obstructive pulmonary disease (COPD) is the fourth leading cause of death in the United States, and COPD morbidity and mortality continue to rise in the United States and around the world (1). In 1999, 124,000 persons died from COPD-related illness in the United States (2) of approximately 16 million persons with the disease (3). Direct US medical costs were approximately \$15 billion in 1993 (4), and COPD is projected to rank fifth in 2020 in the

worldwide burden of disease according to the World Bank and the World Health Organization (5). Despite the enormity of the problem, there are few strategies available for the prevention of COPD in the general population to supplement smoking prevention and cessation.

Although smoking is the primary cause of COPD, smoking does not explain all of the variance in the development of COPD. Only about 15–20 percent of smokers develop symptomatic COPD (6), and approximately 10 percent of COPD-related mortality occurs in persons with no history of cigarette smoking (7). Given that pulmonary function declines with age (8), smoking cessation alone may not be adequate to protect aging persons from developing COPD.

Ongoing prospective cohort studies such as the Nurses' Health Study (9) and the Health Professionals' Follow-up Study (10) potentially provide opportunities to examine nutritional and other lifestyle risk factors for COPD. These two studies have collected large quantities of data on modifiable risk factors biennially over the last 25 years. However, they include large numbers of subjects dispersed throughout the United States and are conducted by mail. Since it was infeasible to conduct personal interviews, physical examinations, and unified, directly measured spirometry on all 121,700 participants in the Nurses' Health Study, we established a questionnaire-based case definition for COPD. All participants in these cohorts have been asked

Received for publication July 26, 2001, and accepted for publication January 11, 2002.

Abbreviations: CI, confidence interval; COPD, chronic obstructive pulmonary disease; FEV<sub>1</sub>, forced expiratory volume in 1 second; FVC, forced vital capacity; NHANES III, Third National Health and Nutrition Examination Survey.

<sup>1</sup>Channing Laboratory, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA.

<sup>2</sup>General Medicine Division, Department of Medicine, Massachusetts General Hospital, Harvard Medical School, Boston, MA.

<sup>3</sup>Department of Epidemiology, Harvard School of Public Health, Boston, MA.

<sup>4</sup>Department of Environmental Health, Harvard School of Public Health, Boston, MA.

<sup>5</sup>Department of Emergency Medicine, Massachusetts General Hospital, Harvard Medical School, Boston, MA.

Reprint requests to Dr. Carlos A. Camargo, Jr., Channing Laboratory, 181 Longwood Avenue, Boston, MA 02115 (e-mail: carlos.camargo@channing.harvard.edu).

about a physician diagnosis of COPD and asthma every 2 years since 1988, and separate supplemental COPD and asthma questionnaires were mailed in 1998. However, it is not known whether these health professionals reliably report a physician diagnosis of COPD.

Therefore, we conducted a validation study of a questionnaire-based case definition of COPD by reviewing the medical records of a stratified, randomly selected sample of participants reporting COPD. We hypothesized that a self-report of a physician diagnosis of COPD from a participant in the Nurses' Health Study on an original and supplemental form is a valid marker of COPD. We also tested the effect of misclassification of COPD in a hypothetical cohort proportional in size to the Nurses' Health Study.

## MATERIALS AND METHODS

### Study sample

In 1976, the Nurses' Health Study enrolled 121,700 married, female registered nurses aged 30–55 years who resided in one of 11 US states and responded to a two-page questionnaire (9). Follow-up questionnaires, mailed to all cohort members every 2 years, inquired about lifestyle exposures and the development of medical conditions. Overall follow-up was greater than 90 percent in 1998. Starting in 1988, all participants were asked biennially about a physician diagnosis of emphysema, a physician diagnosis of chronic bronchitis, and a physician diagnosis of asthma.

### Follow-up and definition of cases

In 1998, a one-page electronically scannable COPD questionnaire was sent to all living nurses who reported a physician diagnosis of emphysema or chronic bronchitis from 1988 to 1996. This supplemental questionnaire requested information confirming a physician diagnosis of emphysema, chronic bronchitis, or COPD; dates of symptom onset and diagnosis; tests performed to confirm the diagnosis; symptoms consis-

tent with a diagnosis of chronic bronchitis; and a concurrent physician diagnosis of asthma. A similar supplemental questionnaire was mailed in the same year to all living nurses who reported a physician diagnosis of asthma from 1988 to 1996. This asthma questionnaire asked about dates of asthma onset, diagnosis, remission, and recurrence; physician diagnosis of COPD; history of atopy; asthma-related hospital admission; and past systemic steroid use. Both questionnaires also included items on recent medication use, respiratory symptoms, health care utilization (hospital admissions, emergency department visits, urgent office visits), and results of spirometry in the preceding year. The supplemental questionnaires were mailed to nonresponders two additional times.

We wanted to use the contemporary clinical definition of COPD: a diagnosis of COPD, emphysema, or chronic bronchitis with evidence of airflow obstruction that is not fully reversible (11–13). Since spirometry data were not available for all participants, we defined self-reported COPD with varying levels of certainty. Using data from the supplemental questionnaire, we categorized participants who self-reported COPD into overlapping groups of “definite,” “possible,” and “probable” cases based on the criteria defined in table 1. Definitions were established independent of smoking status. Cases were excluded if the supplemental questionnaire showed a report of normal spirometry or comorbid pulmonary disease other than asthma. Since COPD is rarely, if ever, diagnosed before age 35 years (11) (in contrast to asthma), cases were also excluded if their reported age at COPD diagnosis was 35 years or less.

### Validation of cases

To assess the validity of these definitions of self-reported COPD, we identified a 10 percent random sample of participants who reported COPD, stratified by reports of diagnostic tests and exclusion criteria. In addition, we sampled a smaller number of nurses who reported only physician-diagnosed chronic bronchitis and who did not report recent symptoms diagnostic of chronic bronchitis.

**TABLE 1. Classification criteria for self-reported chronic obstructive pulmonary disease in the Nurses' Health Study, United States, 1976–2000**

Criterion	Definition
Definite	Report of $FEV_1^*$ < 80% predicted or $FEV_1/FVC^*$ < 70% in the preceding year on the supplemental (1998) questionnaire <i>and</i> meeting the criteria for a “possible” case
Probable	Report of diagnostic test at diagnosis (pulmonary function testing, chest radiograph, or chest computed tomography) <i>and</i> meeting the criteria for a “possible” case
Possible	Affirmative response to physician-diagnosed chronic bronchitis or emphysema on the original (1988–1996) questionnaire <i>and</i> affirmative response to one of the following on the supplemental (1998) questionnaire: physician-diagnosed chronic bronchitis† physician-diagnosed emphysema, physician-diagnosed COPD*
Exclusion criteria	Report of normal $FEV_1$ on the supplemental (1998) questionnaire, age at diagnosis $\leq 35$ years, or comorbid pulmonary disease other than asthma

\*  $FEV_1$ , forced expiratory volume in 1 second; FVC, forced vital capacity; COPD, chronic obstructive pulmonary disease.

† For a physician diagnosis of chronic bronchitis, report of symptoms of productive cough for  $\geq 3$  months in the preceding  $\geq 2$  years was required on the supplemental (1998) questionnaire.

We sent these participants as many as five letters requesting authorization to release and review their medical records in 1999. For participants who provided authorization, we requested medical records consisting of two pulmonary function tests, two chest films, one chest computed tomography scan, two office or emergency department visits for COPD or asthma, and one hospital discharge summary regarding COPD or asthma. As many as three requests were sent to participants' physicians and hospitals. A trained physician (R. G. B.) reviewed all medical records, blinded to self-reported outcomes. Data were extracted using a standardized extraction form, with particular attention to spirometry records and admission notes.

Self-reported cases of disease were confirmed as COPD if spirometry was consistent with a diagnosis of COPD, if the radiologist's interpretation of a chest film or computed tomography scan included COPD or emphysema, or if an attending physician documented a diagnosis of COPD or emphysema in the medical record. Spirometry was classified according to Global Initiative for Chronic Obstructive Lung Disease (GOLD) guidelines, which define moderate-to-severe COPD as forced expiratory volume in 1 second ( $FEV_1$ )/forced vital capacity (FVC) < 0.70, with  $FEV_1$  < 80 percent predicted (13). Documentation of  $FEV_1 \geq 80$  percent predicted or  $FEV_1/FVC \geq 0.70$  generally excluded validation of self-reported COPD. Exceptions were made on an individual basis for participants for whom there was repeated physician documentation of COPD years after a spirometry report showing an  $FEV_1$  value of 80–99 percent predicted by nomogram or  $FEV_1/FVC \geq 0.70$ . GOLD guidelines classify mild COPD as  $FEV_1/FVC < 0.70$  and  $FEV_1 \geq 80$  percent predicted (13).  $FEV_1 \geq 100$  percent predicted excluded validation of self-reported COPD in all instances. A medical record notation of chronic bronchitis was considered inadequate to confirm a self-reported case in the absence of spirometry records meeting diagnostic criteria. Medical records were also reviewed systematically for alternative pulmonary diagnoses and for a physician diagnosis of asthma.

### Statistical analysis

The proportion of cases validated was calculated as confirmed cases divided by reported cases. This proportion corresponds to the positive predictive value of the questionnaire-based definition.

Because reporting of COPD on the questionnaire may be misclassified, we estimated the impact of any potential bias on relative risk estimates in a hypothetical cohort. We performed a data simulation with a hypothetical cohort of 40,000 subjects, half of whom were "exposed" (e.g., high fish oil intake) and half "unexposed" (e.g., low fish oil intake). We specified 1,000 questionnaire-based cases of COPD in this hypothetical cohort, approximating the number of prevalent cases of "possible" COPD reported in the highest and lowest quintiles of a given exposure in the Nurses' Health Study. We distributed these potentially misclassified cases according to possible observed relative risk values varying from 0.4 to 2.5 and assumed that misclassification was independent of exposure.

We then calculated the number of correctly classified cases (i.e., if misclassification were absent) using the following formula (14):  $A = (A^* - Fp \times N)/(Se + Sp - 1)$ , where  $A$  is the number of correctly classified cases of COPD,  $A^*$  is the number of questionnaire-based cases of COPD,  $Fp$  is the false-positive probability ( $1 - \text{sensitivity}$ ),  $N$  is the number of participants within a stratum, and  $Se$  is the sensitivity and  $Sp$  the specificity of a questionnaire-based report of COPD. Unbiased relative risks and 95 percent confidence intervals were calculated from these correctly classified cases.

The sensitivity and specificity of a report of COPD were calculated from the number of true positives and false positives observed in the validation set and by assuming a false-negative probability of 0.01 among a representative 10 percent random sample of the overall cohort. The assumption implies that 1 percent of the cohort had COPD and did not report it. Sensitivity analyses were performed to check this assumption.

Associations between proportions were tested with the  $\chi^2$  test. Student's  $t$  test and one-way analysis of variance were used to compare means of normally distributed continuous variables. Two-tailed  $p < 0.05$  was defined as statistically significant. SAS 7.0 software (SAS Institute, Inc., Cary, North Carolina) was used for analyses.

## RESULTS

Between 1988 and 1996, 4,852 nurses reported a physician diagnosis of COPD in the Nurses' Health Study. Of these nurses, 1,135 reported chronic bronchitis alone without diagnostic symptoms (refer to Materials and Methods), and an additional 927 met exclusion criteria, that is, a report of normal spirometry, age at diagnosis less than or equal to 35 years, or comorbid pulmonary disease other than asthma. Therefore, 2,790 nurses met the case definition of self-reported COPD.

### Response to medical record requests

We randomly selected 422 participants with self-reported COPD to achieve a 10 percent sample, and we sampled an additional 171 persons who reported chronic bronchitis alone without diagnostic symptoms. Thirteen participants who reported COPD and five in the latter group died in the interim; among the remaining 575 persons, 510 responded (89 percent). Of the respondents, 50 reconfirmed the diagnosis but refused access to their medical records, 14 neither confirmed nor denied the diagnosis and refused further participation, and 32 denied both the diagnosis and access to their medical records. Sufficient records were obtained to complete the medical record review for 376 of 414 (91 percent) participants who allowed access to their records. Therefore, medical records were available for 273 (65 percent) of self-reported cases of COPD and 103 (60 percent) persons with chronic bronchitis alone without diagnostic symptoms.

Included participants were similar to women for whom data were incomplete with respect to age and year of diag-

**TABLE 2. Proportion of cases confirmed\* among participants with a self-reported physician diagnosis of chronic obstructive pulmonary disease in the Nurses' Health Study, United States, 1976–2000**

Criterion†	Self-report of COPD‡		Method of confirmation§			FEV <sub>1</sub> %‡ predicted¶ (mean (standard deviation))
	Records obtained (no.)	Cases confirmed in medical record (%)	Pulmonary function test result (%)	Chest radiograph or computed tomography (%)	Physician notation (%)	
Definite	73	86	81	51	60	45 (18)
Probable	218	80	71	50	64	50 (19)
Possible	273	78	67	48	67	50 (19)

\* Confirmed using medical records and uniform diagnostic criteria.

† Refer to table 1 for definitions.

‡ COPD, chronic obstructive pulmonary disease; FEV<sub>1</sub>, forced expiratory volume in 1 second.

§ All three elements were not available for some participants.

¶ Among participants with pulmonary function test reports.

nosis of COPD. However, the former were more likely to have “definite” COPD according to the questionnaire criteria ( $p = 0.04$ ) and to report a concurrent physician diagnosis of asthma ( $p < 0.001$ ).

### Validation

Overall, 78 percent of self-reported cases of COPD fulfilling “possible” criteria were confirmed by medical record review (table 2). The proportion confirmed increased for questionnaire-based categories of “possible” to “probable” to 86 percent for “definite” COPD ( $p$  for trend = 0.02). Participants who reported “definite” COPD were more likely to be confirmed to have COPD by spirometry ( $p < 0.001$ ) than were other participants; mean FEV<sub>1</sub> percent predicted was marginally lower among these participants than among all confirmed cases for whom pulmonary function test data were available ( $p = 0.05$ ). Restricting the analysis to participants who reported a new diagnosis of COPD after 1988 resulted in confirmed proportions of 0.83, 0.84, and 0.90 for “possible,” “probable,” and “definite” COPD, respectively. Physician diagnosis alone confirmed 16 percent of “possible” cases, 13 percent of “probable” cases, and 8 percent of “definite” cases.

Women could report emphysema, chronic bronchitis, and COPD on the supplemental questionnaire. Thirty-six women in the validation sample reported all three, and disease in all 36 of them (100 percent) was confirmed by medical record review. Otherwise, the highest proportion of confirmed cases of disease occurred among women who reported emphysema, followed by COPD, then chronic bronchitis (table 3). Disease in women who reported chronic bronchitis—but not emphysema or COPD—was confirmed 40 percent of the time.

### Overlap with asthma

Fifty-nine percent of participants in the overall cohort who reported COPD also reported a physician diagnosis of asthma. Compared with women who reported only COPD, women who reported both asthma and COPD were less likely to be confirmed to have COPD by medical record review (table 4,  $p < 0.0001$ ). A physician diagnosis of asthma was found in 51 percent of the medical records of women who reported both diseases. On the other hand, asthma was noted in the medical records of 6 percent of women who reported only COPD.

**TABLE 3. Proportion confirmed\* among participants with self-reported physician diagnosis of chronic obstructive pulmonary disease in the Nurses' Health Study, according to self-report of emphysema, chronic bronchitis, and chronic obstructive pulmonary disease, United States, 1976–2000**

Criterion†	Self-report of					
	Emphysema		Chronic bronchitis		COPD‡	
	Records obtained (no.)	Cases confirmed in medical record (%)	Records obtained (no.)	Cases confirmed in medical record (%)	Records obtained (no.)	Cases confirmed in medical record (%)
Definite	39	92	43	84	42	93
Probable	114	91	116	79	123	85
Possible	145	89	139	79	156	81

\* Confirmed using medical records and uniform diagnostic criteria.

† Refer to table 1 for definitions.

‡ COPD, chronic obstructive pulmonary disease.

**TABLE 4. Proportion of cases confirmed\* among participants with a self-reported physician diagnosis of chronic obstructive pulmonary disease with and without a physician diagnosis of asthma in the Nurses' Health Study, United States, 1976–2000**

Criterion†	Self-report of			
	COPD‡ only		Asthma and COPD	
	Records obtained (no.)	Cases confirmed in medical record (%)	Records obtained (no.)	Cases confirmed in medical record (%)
Definite COPD	24	96	49	82
Probable COPD	94	88	124	73
Possible COPD	125	86	148	70

\* Confirmed using medical records and uniform diagnostic criteria.

† Refer to table 1 for definitions.

‡ COPD, chronic obstructive pulmonary disease.

The order of diagnosis of COPD and asthma was examined. Proportions of confirmed COPD were similar for women who reported asthma prior to ( $n = 56$ ), concurrent with ( $n = 87$ ), and following ( $n = 27$ ) a diagnosis of COPD (71, 71, and 59 percent, respectively;  $p = 0.34$ ). Similar proportions of these women had a physician diagnosis of asthma in their medical records (57, 49, and 48 percent, respectively;  $p = 0.61$ ). Women who reported both COPD and asthma were equally likely to have partial reversibility to bronchodilators on spirometry records as women who reported only COPD (23 and 25 percent, respectively;  $p = 0.70$ ). Order of diagnosis of asthma and COPD was not related to evidence of partial reversibility ( $p = 0.66$ ).

#### Unconfirmed cases

Sixty-one self-reports of COPD were not confirmed by medical record review. Fifty-seven percent of these 61 women's medical records included a diagnosis of asthma, and 20 percent had an alternative pulmonary diagnosis such as sarcoidosis or pulmonary embolism.

Thirty-two women who refused access to their medical records also denied that they had previously reported a diagnosis of COPD. However, three women in this group had previously reported a hospitalization for COPD, and 50 percent had reported an emergency room or urgent office visit for COPD in 1998. Sixty percent reported taking medication for COPD during the week before the 1998 supplemental questionnaire was administered, and the median number of pack-years of smoking at the beginning of follow-up (1988) was 40.

#### Effect of misclassification

In the data simulation, nondifferential misclassification of COPD of the magnitude found in the validation sample biased relative risks toward the null value, as shown in table 5. For example, if the relative risk estimated from questionnaire-based cases of COPD was 0.50 (95 percent confidence interval (CI): 0.44, 0.57), the unbiased relative risk estimated from correctly classified cases of COPD was 0.39 (95 percent CI: 0.34, 0.44) after correction for misclassification (table 5). As

**TABLE 5. Data simulation comparing relative risk estimates obtained by using the questionnaire-based approach with those obtained by using a correctly classified approach to ascertain chronic obstructive pulmonary disease in a hypothetical cohort\***

Questionnaire based				Classified correctly			
COPD cases among 20,000 subjects		Observed relative risk	95% confidence interval	COPD cases among 20,000 subjects		Unbiased relative risk	95% confidence interval
Exposed	Unexposed			Exposed	Unexposed		
286	714	0.40	0.35, 0.46	242	879	0.28	0.24, 0.32
333	667	0.50	0.44, 0.57	313	808	0.39	0.34, 0.44
400	600	0.67	0.59, 0.76	412	709	0.58	0.52, 0.66
500	500	1.00	0.88, 1.13	561	561	1.00	0.89, 1.12
600	400	1.50	1.32, 1.70	709	412	1.72	1.53, 1.94
667	333	2.00	1.76, 2.28	808	313	2.58	2.27, 2.93
714	286	2.50	2.18, 2.86	879	242	3.63	3.15, 4.17

\* The hypothetical cohort consisted of 20,000 "exposed" and 20,000 "unexposed" subjects, 1,000 of whom reported a physician diagnosis of chronic obstructive pulmonary disease (COPD) on a questionnaire. A variety of observed relative risks are shown. The correctly classified cases are estimated from the questionnaire-based cases (refer to Materials and Methods); unbiased relative risks show the corresponding relative risks if no misclassification had been present. Correction is based on a positive predictive value of 0.78 and assumes a false-positive rate of 0.01 (i.e., 1% of the entire cohort has COPD that is not reported) and that misclassification from the questionnaire is random with respect to the exposure.

anticipated, the same pattern was found for harmful exposures: a questionnaire-based relative risk of 2.00 (95 percent CI: 1.76, 2.28) suggested an unbiased relative risk of 2.58 (95 percent CI: 2.27, 2.93).

These calculations assumed a false-negative probability of 0.01; that is, 1 percent of the cohort had COPD and did not report it. Sensitivity analyses to check this assumption showed that it was highly robust and that changes in false-negative probabilities had little impact on relative risks compared with changes in false-positive probabilities. Values of false-negative probabilities ranging from 0 to 0.02 produced results that were within two decimal places of the original estimate of 2.58 (95 percent CI: 2.27, 2.93). Even extreme values of false-negative probability produced similar bias. For example, a false-negative probability of 0.10 (i.e., 10 percent of the cohort had COPD and did not report it) produced an unbiased relative risk of 2.67 (95 percent CI: 2.52, 2.85) for a questionnaire-based relative risk of 2.00. On the other hand, changes in false-positive probabilities produced widely diverging relative risks: the corrected relative risks for a false-positive probability of 0 was 2.00 (95 percent CI: 1.79, 2.23) and for a false-positive probability of 0.015 was 11.14 (95 percent CI: 8.59, 14.45). In contrast, risk differences were strongly affected by changes in false-negative probabilities, whereas risk differences were essentially unaffected by changes in false-positive probabilities (data not shown).

## DISCUSSION

Of a random sample of women self-reporting physician-diagnosed COPD in the Nurses' Health Study, 78 percent of the cases of disease were confirmed by medical record review using uniform diagnostic criteria. Plans are under way to analyze the overall cohort using incident "probable" COPD as the primary study endpoint, for which 84 percent were confirmed. This finding suggests that a questionnaire-based approach to the epidemiologic study of COPD is practical among middle-aged health professionals and that a self-report of COPD from these health professionals is a valid marker of medical record evidence of COPD.

Data from the Third National Health and Nutrition Examination Survey (NHANES III), a representative sample of the general US population, suggest that our validation results are not generalizable to the overall population. In NHANES III, the positive predictive value of a physician diagnosis of COPD for low lung function (according to American Thoracic Society criteria) can be calculated as 0.33 (15). In contrast, the positive predictive value in our study was 0.78 for medical record documentation of COPD overall and 0.71 for low lung function (excluding participants for whom lung function data were missing, as was done in the NHANES III analysis). These proportions improved with more refined, questionnaire-based definitions of COPD.

The difference in these results is likely due to our repeated questionnaire administration (i.e., affirmative responses were required on two questionnaires separated by 2–10 years), exclusion of COPD diagnoses prior to age 35

years, the older average age of participants, the requirement for symptoms diagnostic of chronic bronchitis if chronic bronchitis alone were reported, and a higher prevalence of COPD. In addition, since a woman is less likely than is a man with an identical medical history to receive a physician diagnosis of COPD (16, 17), women have a higher probability of having COPD if labeled with the diagnosis. Finally, health professionals are more likely than nonhealth professionals to reliably report their medical history.

Although the overall proportion of women confirmed to have COPD in our study was high, it varied by "type" of COPD. The lower proportion for chronic bronchitis (with symptoms) is probably due in part to our validation process; because we decided to adhere to the contemporary definition of COPD (11–13), we concentrated on medical record evidence of obstruction on spirometry. Therefore, women who reported chronic bronchitis and had a documented diagnosis of chronic bronchitis on their chart were not considered to have COPD unless there was also evidence of airway obstruction. Many women whose disease was not validated probably had simple, nonobstructive chronic bronchitis.

There was a large overlap of asthma and COPD in our data, similar to population-based studies of physician diagnoses of obstructive airways disease (15, 16). Although the epidemiologic definition of asthma is unclear (18), a single questionnaire item on prior diagnosis of asthma has been shown to effectively separate patients with asthmatic bronchitis from patients with purely smoking-related COPD (8). The higher proportion of cases confirmed for COPD only compared with asthma and COPD suggests a similar pattern in our data. However, stratification by order of diagnosis in the group that reported both diseases was generally unilluminating.

Lacking FEV<sub>1</sub> data on all 121,700 members of this cohort, we adopted a diagnosis-based case definition of COPD among these nurses. Use of a physician diagnosis of COPD as an outcome has been criticized in the past as subject to bias and misclassification (19). However, common biases that might apply selectively to cases and not to other members of the Nurses' Health Study are reduced by the prospective cohort design, standardized questionnaires, consistent questionnaire administration, and very high follow-up rates.

Misclassification of the outcome is unavoidable, given the questionnaire-based design, but the bias that it produces may be predictable. Nondifferential misclassification will generally bias relative risks toward the null value (20). If the false-positive probability is known—from a validation study, for example—and if one assumes that misclassification is nondifferential and that exposure is classified correctly, then misclassification can be corrected (21). Given the false-positive probability calculated from this validation study and the assumptions above, it is possible to estimate corrected, unbiased relative risks, as we showed in the data simulation.

COPD is underdiagnosed in the general US population (15, 22); therefore, even with perfect reporting, it is likely that a sizable number of undiagnosed cases would be

missed. Spirometry data for a random sample of participants in the overall cohort would be one way to quantify the number of undetected cases of COPD (false negatives), but this approach was impractical given the geographic dispersion of participants throughout the United States and contact by mail only.

Such measurement of false negatives is not necessary for correct estimation of relative risks as long as misclassification is nondifferential (21). As we showed in the sensitivity analysis, bias of relative risks from nondifferential disease misclassification is driven by false positives, not false negatives. Assuming extreme and unlikely values of false negatives, such as 0 or 10 percent of the cohort having unreported COPD, produced trivial changes in relative risks. When person-years is used as the unit of measurement, false negatives that are balanced across exposure groups have no effect on the relative risk (14). On the other hand, false negatives do bias the risk difference and are required for the calculation of sensitivity and specificity (23). Since we were primarily interested in relative risks in the person-years data, results from this study are sufficient, and it is unnecessary to sample noncases to find all false negatives.

In summary, the proportion of validated self-reported cases of COPD among participants in the Nurses' Health Study was substantial and sufficient to allow accurate estimation of relative risks. The use of self-reported COPD from health professionals provides an opportunity to explore hypotheses related to nutritional and lifestyle risk factors in larger population groups than would be feasible if directly measured spirometry were required to rule in or rule out COPD. With these methods, it is hoped that additional risk factors for COPD will be found, modification of which may complement smoking cessation.

## ACKNOWLEDGMENTS

Funding was provided by grants HL-07427, PE-11001, HL-03533, and CA-87969 from the National Institutes of Health.

The authors thank Karen Corsano, Gary Chase, and Barbara Egan for their invaluable assistance in implementing the study and Drs. Meir J. Stampfer and Scott T. Weiss for their helpful comments on the manuscript.

## REFERENCES

- Petty TL, Weinmann GG. Building a national strategy for the prevention and management of and research in chronic obstructive pulmonary disease. National Heart, Lung, and Blood Institute Workshop Summary. *JAMA* 1997;277:246-53.
- Hoyert DL, Arias E, Smith BL, et al. Death: final data for 1999. National vital statistics reports 2001. Hyattsville, MD: National Center for Health Statistics, 2001. (US Department of

- Health and Human Services publication PHS 2001-1120).
- Adams PF, Hendershot GE, Marano MA. Current estimates from the National Health Interview Survey, 1996. Hyattsville, MD: National Center for Health Statistics, 1999. (US Department of Health and Human Services publication PHS 99-1528).
- Strategies in preserving lung health and preventing COPD and associated diseases. The National Lung Health Education Program (NLHEP). *Chest* 1998;113(suppl):123S-163S.
- Murray CJL, Lopez AD. Evidence-based health policy—lessons from the Global Burden of Disease Study. *Science* 1996;274:740-3.
- Snider G. Chronic obstructive pulmonary disease: risk factors, pathophysiology, and pathogenesis. *Annu Rev Med* 1989;40:411-29.
- Higgins M. Risk factors associated with chronic obstructive lung disease. *Ann N Y Acad Sci* 1991;624:7-17.
- Burrows B, Knudson RJ, Cline MG, et al. A reexamination of risk factors for ventilatory impairment. *Am Rev Respir Dis* 1988;138:829-36.
- Colditz GA. The Nurses' Health Study: a cohort of US women followed since 1976. *J Am Med Women's Assoc* 1995;50:40-4.
- Rimm EB, Giovannucci EL, Stampfer MJ, et al. A prospective study of alcohol consumption and the risk of coronary disease in men. *Lancet* 1991;338:464-8.
- Standards for the diagnosis and care of patients with chronic obstructive pulmonary disease. American Thoracic Society. *Am J Respir Crit Care Med* 1995;152(suppl):S77-S121.
- BTS guidelines for the management of chronic obstructive pulmonary disease. The COPD Guidelines Group of the Standards of Care Committee of the BTS. *Thorax* 1997;52(suppl 5):S1-S28.
- Pauwels RA, Buist AS, Calverley PMA, et al. Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease. NHLBI/WHO Global Initiative for Chronic Obstructive Lung Disease (GOLD) Workshop Summary. *Am J Respir Crit Care Med* 2001;163:1256-76.
- Greenland S. Basic methods for sensitivity analysis and external adjustment. In: Rothman KJ, Greenland S, eds. *Modern epidemiology*. Philadelphia, PA: Lippincott Williams & Wilkins, 1998:343-57.
- Mannino DM, Gagnon RC, Petty TL, et al. Obstructive lung disease and low lung function in adults in the United States: data from the National Health and Nutrition Examination Survey, 1988-1994. *Arch Intern Med* 2000;160:1683-9.
- Dodge R, Cline MG, Burrows B. Comparisons of asthma, emphysema, and chronic bronchitis diagnoses in a general population sample. *Am Rev Respir Dis* 1986;133:981-6.
- Chapman KR, Tashkin DP, Pye DJ. Gender bias in the diagnosis of COPD. *Chest* 2001;119:1691-5.
- Woolcock A. Epidemiologic methods for measuring prevalence of asthma. *Chest* 1987;91(suppl 1):89S-92S.
- Rijcken B, Schouten JP, Rosner B, et al. Is it useful to distinguish between asthma and chronic obstructive pulmonary disease in respiratory epidemiology? *Am Rev Respir Dis* 1991;143:1456-7.
- Copeland KT, Checkoway H, McMichael AJ, et al. Bias due to misclassification in the estimation of relative risk. *Am J Epidemiol* 1977;105:488-95.
- Brenner H, Gefeller O. Use of the positive predictive value to correct for disease misclassification in epidemiologic studies. *Am J Epidemiol* 1993;138:1007-15.
- Stang P, Lydick E, Silberman C, et al. The prevalence of COPD: using smoking rates to estimate disease frequency in the general population. *Chest* 2000;117(suppl 2):354S-9S.
- Rodgers A, MacMahon S. Systemic underestimation of treatment effects as a result of diagnostic test inaccuracy: implications for the interpretation and design of thromboprophylaxis trials. *Thromb Haemost* 1995;73:167-71.