

# Validity of chronic obstructive pulmonary disease diagnoses in a large administrative database

Yves Lacasse MD MSc FRCP<sup>1</sup>, Jean-Marc Daigle MSc<sup>2</sup>, Sylvie Martin MSc<sup>1</sup>, François Maltais MD<sup>1</sup>

Y Lacasse, J-M Daigle, S Martin, F Maltais. Validity of chronic obstructive pulmonary disease diagnoses in a large administrative database. *Can Respir J* 2012;19(2):e5-e9.

**BACKGROUND:** Administrative databases are often used for research purposes, with minimal attention devoted to the validity of the included diagnoses.

**AIMS:** To determine whether the principal diagnoses of chronic obstructive pulmonary disease (COPD) made in hospitalized patients and recorded in a large administrative database are valid.

**METHODS:** The medical charts of 1221 patients hospitalized in 40 acute care centres in Quebec and discharged between April 1, 2003 and March 31, 2004, with a principal discharge diagnosis of COPD (*International Classification of Diseases, Ninth Revision* codes 491, 492 or 496) were reviewed. The diagnosis of COPD was independently adjudicated by two pulmonologists using clinical history (including smoking status) and spirometry. The primary outcome measure was the positive predictive value (PPV) of the database for the diagnosis of COPD (ie, the proportion of patients with an accurate diagnosis of COPD corroborated by clinical history and spirometry).

**RESULTS:** The diagnosis of COPD was validated in 616 patients (PPV 50.4% [95% CI 47.7% to 53.3%]), with 372 patients (30.5%) classified as 'indeterminate'. Older age and female sex were associated with a lower probability of an accurate diagnosis of COPD. Hospitalization in a teaching institution was associated with a twofold increase in the probability of a correct diagnosis.

**CONCLUSIONS:** The results support the routine ascertainment of the validity of diagnoses before using administrative databases in clinical and health services research.

**Key Words:** Chronic obstructive lung disease; Database; Epidemiology; Validity

Health authorities (often the payers of health care) create and maintain administrative databases by compiling claims data sets. Claims data include the patient diagnosis that motivated the provision of services and the charges paid for the services provided. Typically, the database includes patient demographics and patient-level data about their use of health care resources. Administrators and health care researchers can access the information in these databases to ascertain resource use, even if it involved several providers and health care centres (1-3). When one payer reimburses all health care provisions, these databases afford the opportunity to conduct large population-based observational studies with minimal referral bias, nonresponse and drop-outs.

Similar to other investigators, we were interested in exploiting such a database for a series of studies that could answer health services questions (eg, utilization or quality of care) and clinical questions related to chronic obstructive pulmonary disease (COPD). Before doing so, we considered the underlying validity of the diagnoses included in the database. The objective of the present study was, therefore, to determine the extent to which the principal diagnoses of COPD made in hospitalized patients and recorded in a large administrative database were valid, ie, corroborated by clinical history (including smoking status) and pulmonary function tests.

## La validité de diagnostics de maladie pulmonaire obstructive chronique dans une grande base de données administrative

**HISTORIQUE :** Les bases de données administratives sont souvent utilisées pour des besoins de recherche, et on porte peu attention à la validité des diagnostics qui y figurent.

**OBJECTIFS :** Déterminer si les principaux diagnostics de maladie pulmonaire obstructive chronique (MPOC), posés chez les patients hospitalisés et saisis dans une grande base de données administrative, sont valides.

**MÉTHODOLOGIE :** Les chercheurs ont analysé les dossiers médicaux de 1 221 patients hospitalisés dans 40 centres de soins aigus du Québec et qui avaient reçu leur congé entre le 1<sup>er</sup> avril 2003 et le 31 mars 2004 et dont le principal diagnostic au congé en était un de MPOC (codes 491, 492 ou 496 de la *Classification internationale des maladies, neuvième révision*). Le diagnostic de MPOC était confirmé de manière indépendante par deux pneumologues, d'après les antécédents cliniques (y compris le tabagisme) et la spirométrie. La principale mesure d'issue était la valeur prédictive positive (VPP) de la base de données pour le diagnostic de MPOC (c.-à-d. la proportion de patients ayant reçu un diagnostic exact de MPOC corroboré par les antécédents cliniques et la spirométrie).

**RÉSULTATS :** Le diagnostic de MPOC a été validé chez 616 patients (VPP 50,4 % [95 % IC 47,7 % à 53,3 %]), et 372 patients (30,5 %) ont été classés comme « indéterminés ». L'âge plus avancé et le sexe féminin s'associaient à une plus faible probabilité de diagnostic pertinent de MPOC. L'hospitalisation dans un établissement d'enseignement doublait la probabilité de bon diagnostic.

**CONCLUSIONS :** Les résultats appuient l'évaluation habituelle de la validité des diagnostics avant d'utiliser les bases de données administratives dans le cadre de recherches sur les services cliniques et de santé.

## METHODS

### Data source

In Quebec, the *Régie de l'assurance-maladie du Québec* (RAMQ) is the government body responsible for universal medical insurance registration for the 7.5 million people that live in the province. Since 1980, the RAMQ has maintained a database (MedEcho) that encrypts hospitalization data (including principal diagnosis at discharge, secondary diagnoses, therapeutic and diagnostic procedures and length of stay) for all patients. The MedEcho database has been used for research purposes on several occasions to describe the epidemiology of diseases (4), or to quantify the risks (5,6) or benefits (7) of medical interventions.

### Sampling (Table 1)

A cohort study was conducted using a sample of acute care hospitals in Quebec. First, 97 acute care hospitals were divided into four quartiles according to the number of discharges per hospital with a principal diagnosis of COPD (*International Classification of Diseases, Ninth Revision* [ICD-9] codes 491, 492 or 496) between April 1, 2003 and March 31, 2004. Within each quartile, 10 centres were randomly selected with a probability proportional to the number of hospital discharges for COPD. In a second step, patients with a principal diagnosis of COPD at discharge during the same period were randomly selected

<sup>1</sup>Centre de recherche, Institut universitaire de cardiologie et de pneumologie de Québec (Hôpital Laval); <sup>2</sup>Direction des systèmes de soins et politiques publiques, Institut national de santé publique du Québec, Québec, Québec

Correspondence: Dr Yves Lacasse, Centre de Pneumologie, Institut universitaire de cardiologie et de pneumologie de Québec (Hôpital Laval), 2725 Chemin Ste-Foy, Québec, Québec G1V 4G5. Telephone 418-656-4747, fax 418-656-4762, e-mail yves.lacasse@med.ulaval.ca

		Cohort of patients	
		COPD +	COPD –
Administrative database	COPD +	a	b
	COPD –	c	d

**Figure 1** Two properties of an administrative database. Positive predictive value: the proportion of patients with a diagnosis of chronic obstructive pulmonary disease (COPD) in the database who truly have COPD (+) ( $a/[a + b]$ ). Sensitivity: the proportion of patients with COPD who are correctly entered in the database as having COPD ( $a/[a + c]$ )

from the database with equal probabilities. Therefore, secondary diagnoses of COPD were not considered. In case of multiple hospitalizations during the study period, only the first was considered. All data were obtained from the RAMQ with permission from the *Commission de l'accès à l'information du Québec*.

### Procedures

A list of health insurance and hospital file numbers was forwarded to the medical records department of each of the 40 randomly selected hospitals. The relevant data were extracted from patient charts by a medical archivist designated by each institution using a pretested data extraction form. The extracted data included the principal diagnosis at discharge, smoking status (current smoker, exsmoker or never smoker), copies of spirometry data obtained during the index hospitalization and before (up to  $n=12$ ), and the summary of the hospitalization, on which the name and specialty of the attending physician was found. The information was anonymized by the archivist before sending it to the study centre.

### 'Gold standard' defining COPD and patient classification

Two pulmonologists (YL and FM) with extensive experience in COPD management independently reviewed the available clinical information and spirometry data to classify patients into three categories: COPD; indeterminate; and alternative diagnosis. The spirometry data were interpreted according to the 2007 Global Initiative for Chronic Obstructive Lung Disease (GOLD) classification (8). To conform with the GOLD definition of COPD, COPD was diagnosed if the available spirometry data demonstrated airflow obstruction (postbronchodilator forced expiratory volume in 1 s [ $FEV_1$ ]/forced vital capacity [FVC] ratio  $<70\%$ ) in an individual with past or current smoking status. In these patients, the best available  $FEV_1$  was noted during the two years that preceded the hospitalization to classify the severity of COPD.

When no spirometry data were available, or when smoking status was unknown, patients were classified as 'indeterminate'. Patients whose spirometry data were uninterpretable were also included in this category. Patients classified as not having COPD (ie, alternative diagnosis) were those for whom COPD could be definitely excluded from the available data. These patients were further classified into two categories: 'codification error' or 'medical misdiagnosis'. Codification errors occurred at the clerical level and resulted from misinterpretation of ICD rules. A typical example would be a patient with cystic fibrosis correctly diagnosed with an exacerbation by the treating physician, with an entry of 'COPD' in MedEcho. Medical misdiagnoses occurred at the physicians' level and included, for example, patients discharged with a diagnosis of COPD despite the absence of airway obstruction on spirometry. In such cases, a specific diagnosis (other than 'not COPD') was not mandatory.

### Statistical analysis

**Sample size:** A priori, it was arbitrarily determined that information in the database would be considered valid if the positive predictive value (PPV) (ie, the proportion of patients with a diagnosis of COPD who

truly had COPD – Figure 1) was at least 85%. Accordingly, a computed sample size of 1224 patients was necessary to obtain a precision of  $\pm 2\%$  of this estimate (9).

**Analysis:** Continuous variables were expressed as means  $\pm$  SDs, whereas categorical data were expressed as percentages. For continuous variables, values were compared by using F tests in ANOVA. Differences in proportions were compared using  $\chi^2$  tests. The concordance between patient classification by the two reviewers (YL and FM) was examined using the Kappa statistic (10). In the main analysis, the proportion (and associated 95% CIs) of patients in each of the three final categories (COPD, indeterminate and alternative diagnosis) was determined. Finally, logistic regression analyses were conducted using patients', hospitals' and physicians' characteristics, with 'COPD' as the dependent variable to identify factors predicting an appropriate diagnosis of COPD. Statistical significance was set at the 0.05 level.

**Sensitivity analysis: COPD in nonsmokers:** In the main analysis, the diagnosis of COPD was accepted only in patients with current or history of smoking status. This criterion was also applied in the most recently published large trials of pharmacological and nonpharmacological interventions in COPD (11-16). However, never smokers may represent a significant proportion of adults with airway obstruction (17). The contribution of never smokers to the total burden of COPD remains uncertain (18). In the sensitivity analysis, the same analyses were repeated, considering abnormal spirometry as the only requirement to diagnose COPD, and accepting the diagnosis of COPD in never smokers.

## RESULTS

### Population

Between April 1, 2003 and March 31, 2004, 11,967 patients were discharged with a principal diagnosis of COPD from 97 acute care hospitals in Quebec. The sample of 1224 patients thus represented 10.2% of the total population. The distributions of discharge volume into quartiles and patient sampling are summarized in Table 1. Three patients were excluded for the following reasons: charts for two patients could not be retrieved, and no information other than the principal diagnosis of 'emphysema' was available in the other. Therefore, 1221 patients were included in the present analysis.

### Validity of diagnoses

The concordance between the two reviewers in their classification of patients was near perfect (Kappa 0.91 [95% CI 0.88 to 0.95]). Limiting the agreement study to the most difficult cases (ie, those for whom both spirometry and smoking status were available [ $n=785$ ]), excellent agreement was found between the reviewers (Kappa 0.85 [95% CI 0.79 to 0.90]). Disagreement was always resolved by consensus.

Of the 1221 patients, the diagnosis of COPD was validated in 616 (PPV 50.4% [95% CI 47.7% to 53.3%]), while 372 patients (30.5% [95% CI 27.9% to 33.1%]) were classified as indeterminate. In the remaining 233 patients (19.1% [95% CI 16.9% to 21.3%]), the diagnosis of COPD could be dismissed based on the available data. In these 233 patients, codification errors accounted for only 20% of the reasons for classifying as 'alternate diagnoses' ( $n=47$ ). The remaining 186 patients represented 'medical misdiagnoses'.

In the sensitivity analysis, at least one spirometry was retrieved in 87 never smokers, 48 of whom had airway obstruction. Therefore, considering abnormal spirometry as the only requirement to diagnose COPD (ie, accepting the diagnosis of COPD in never smokers), the PPV of the database was (616 $\pm$ 48)/1221 (54.4% [95% CI 51.5% to 57.2%]).

### COPD characteristics

Patient characteristics according to the three diagnostic categories are presented in Table 2. With the exception of length of stay, significant differences were found in all characteristics considered, including hospital mortality. Among the 616 patients who were properly diagnosed with COPD, 479 (77.8%) had undergone at least one spirometry during the index hospitalization or within the preceding two years.

**TABLE 1**  
Total population and patient sampling

	Quartile				Total
	1	2	3	4	
Sampling frame					
Acute care hospitals in Quebec	24	23	25	25	97
Discharges per hospital with a principal diagnosis of COPD*	<69	69–128	129–260	≥261	–
Patients with a principal diagnosis of COPD*	682	1661	2999	6625	11,967
Sample					
Hospitals	10	10	10	10	40
Patients	67	167	303	687	1224

Data presented as n. \*Between April 1, 2003 and March 31, 2004. COPD Chronic obstructive pulmonary disease

**TABLE 2**  
Patient characteristics

Characteristic	Total (n=1221)	COPD (n=616)	Indeterminate (n=372)	Not COPD (n=233)	P
Age*, years, mean ± SD	73.1±12.2	72.3±9.5	75.5±10.5	71.3±18.8	<0.0001
Male patients	52.2	55.7	53.0	41.6	0.0012
Length of stay*, days, mean ± SD	11.0±17.6	11.6±22.0	10.4±12.7	10.2±9.6	0.46
Smoking status					
Smoker	81.9	100.0	82.3	33.5	<0.0001
Indeterminate	7.5	0.0	10.5	22.3	
Never smoker	10.6	0.0	7.2	44.2	
Spirometry					
None	23.3	0.0	69.4	9.7	<0.0001
1	19.5	22.4	10.8	25.8	
2	12.1	13.7	4.0	21.0	
3	8.8	13.3	4.0	4.8	
≥4	36.5	50.6	11.8	38.7	
Hospital mortality	5.1	4.7	7.5	2.2	0.0113

Data presented as % unless otherwise indicated. \*Continuous variable. COPD Chronic obstructive pulmonary disease

According to the GOLD classification, 12 (2.5%) were classified as GOLD I COPD ( $FEV_1 \geq 80\%$  predicted), 158 (33.0%) as GOLD II ( $50\% \leq FEV_1 < 80\%$  predicted), 206 (43.0%) as GOLD III ( $30\% \leq FEV_1 < 50\%$  predicted) and 103 (21.5%) as GOLD IV ( $FEV_1 < 30\%$  predicted).

### Predictors of a proper diagnosis of COPD

In the univariate analyses, four significant predictors of a proper diagnosis of COPD were identified (Table 3). Older age was associated with a lower probability of a proper diagnosis of COPD. Men were 33% more likely to be discharged with an accurate diagnosis of COPD than women. Pulmonologists were better than any other clinicians to diagnose the disease. Finally, being hospitalized in a teaching hospital was associated with a twofold increase in the probability of a correct diagnosis. Three factors (age, sex and type of hospital) remained significant in the multivariable analysis, with ORs very similar to those obtained in the univariate analyses (data not shown).

### DISCUSSION

Although the definition of COPD we applied was less stringent than that used in most prospective academic or pharmaceutical studies, the diagnosis of COPD could not be ascertained in almost one-half of our large cohort. Considering abnormal spirometry as the only requirement to diagnose COPD (ie, accepting the diagnosis of COPD in never smokers) did not influence the results of the study. We interpret the significant differences in mortality among the three groups (COPD, indeterminate and not COPD) as an indication that the three groups truly represent different populations. Our results, therefore, raise serious questions and concerns regarding the use of administrative databases at the clinical, epidemiological and research levels.

On clinical grounds, our study demonstrated that spirometry is underused because no spirometry data were available in almost one-quarter of our cohort. We classified those patients as 'indeterminate'.

**TABLE 3**  
Univariate analysis of predictors for chronic obstructive pulmonary disease (COPD)

Predictor	OR (95% CI)	P	P (global)
Age, years			
<55	0.30 (0.17–0.54)	<0.0001	
55–64 (reference)	1.00		
65–74	0.69 (0.46–1.02)	0.0636	
≥75	0.42 (0.29–0.61)	<0.0001	<0.0001
Male sex	1.33 (1.06–1.66)	0.0133	0.0133
Length of stay, days			
≤7 (reference)	1.00		
>7	1.05 (0.84–1.32)	0.65	0.65
Attending physician			
Pulmonologist (reference)	1.00		
General practitioner	0.36 (0.25–0.52)	<0.0001	
Other specialist	0.40 (0.24–0.66)	0.0004	<0.0001
Hospital volume*			
10–68 (reference)	1.00		
69–128	1.21 (0.69–2.14)	0.51	
129–260	1.17 (0.69–2.00)	0.56	
≥261	1.02 (0.62–1.70)	0.93	0.64
Hospital			
Nonteaching (reference)	1.00		
Teaching	1.87 (1.41–2.48)	<0.0001	<0.0001

\*Discharges (n) per hospital with a principal diagnosis of COPD

We understand that many clinicians may correctly diagnose COPD from clinical history and physical examination, typical radiological findings of emphysema or the results of arterial blood gas analysis,

especially when hypercapnia is found. However, when spirometry data were unavailable, clinicians' overall impressions predicted airflow limitation only moderately well (19). The proportion of correct diagnoses of COPD among patients labelled 'indeterminate' is, therefore, unknown. Only spirometry can confirm airflow obstruction and COPD (8). The third United States National Health and Nutrition Survey (NHANES III), and the more recent Burden of Obstructive Lung Disease (BOLD) Study emphasized that spirometry is not used widely in clinical practice (20,21). In both studies, this situation translated into a high rate of undiagnosed obstructive lung disease. In the hospital setting, our results suggest that the absence of spirometry has a paradoxical effect: more patients are inaccurately diagnosed with COPD. Furthermore, our finding that women were more often misdiagnosed than men confirms the sex bias in the diagnosis of COPD (22,23). In addition, false inaccurate diagnoses of COPD are likely followed by inappropriate long-term pharmacological therapy, with possible clinical and economic consequences (24).

At the epidemiological level, our results raise serious concerns regarding the validity of statistics derived from hospital records, which ultimately contribute to national statistics. At the research level, the present study challenges epidemiologists and clinical researchers who wish to use administrative data to devote attention to their validity. In studies that included a systematic review of discharge coding accuracy, the proportion of hospital episode records that had the same code as that assigned by an independent review of the discharge summary or medical record ranged from 53% to 100% (25). The authors noted that coding accuracy rates were higher for high-prevalence conditions, which contradicts our results. COPD was not considered in any of the studies that met the inclusion criteria of this systematic review. The authors concluded that policy makers, planners and researchers need to recognize and account for the degree of inaccuracy in routine hospital information statistics.

Based on the results of that systematic review and our findings, we would certainly dismiss the results of any research that used data from an administrative database without previous comprehensive validation, especially when the results bear on clinical practice. Also important is that the validity of databases is diagnosis specific (26). For example, validity for the diagnosis of diabetes by no means implies validity for the diagnosis of COPD. In addition, the decision to use an administrative database should be made on validity thresholds determined a priori. We arbitrarily determined a priori that we would consider the database to be valid if the PPV was at least 85%. Our study demonstrated that the MedEcho database does not meet this criterion of quality. Consequently, we abandoned our intent to use the database for subsequent studies. This contrasts with the study by Rutten et al (27), which examined beta-blockers in COPD that used a general practitioners network database, in which only approximately 70% of the diagnoses of COPD conformed to the GOLD criteria. We would argue that their database did not meet the minimal criterion of quality for use in clinical research. Finally, validation studies must be regularly updated if databases are to be used repeatedly over time.

We acknowledge that administrative databases can be used effectively in certain circumstances (3). For instance, McKnight et al (28) verified the reliability with which another of RAMQ's databases recorded clinical diagnoses of COPD and asthma. They found excellent agreement between the physician service claims files in the database and the medical records of patients with well-characterized diagnoses of COPD and asthma. Such a design tests the sensitivity of a database, ie, the extent to which a given diagnosis appears in the database when patients with this diagnosis seek medical attention (ie,  $a/[a + c]$ ) in Figure 1). High sensitivity would support the use of the database, for instance, to study health care utilization in a cohort of patients known to have COPD or asthma. We previously referred to this property of an administrative database as its "validity of recording" (29). This contrasts with many researchers' improper use of administrative databases to identify large cohorts of patients without ascertainment of clinical diagnosis. Such a design would require that the

administrative database had a high 'PPV' ( $a/[a + b]$  in Figure 1), a property we previously referred to as the "validity of diagnosis" (29).

A major strength of our study is that we relied on medical history (including smoking status) and spirometry to validate the diagnosis of COPD. The most obvious limitation of our study is that the diagnosis of COPD may have been verified elsewhere, without objective documentation by spirometry in the patients' hospital file. We classified these patients as 'indeterminate'. However, we acknowledge that this situation occurred in only a small proportion of patients. The primary reason for this is that spirometry is underused in Quebec because there is little monetary incentive for family physicians to perform it in their office. As a result, during the study period (April 1, 2003 to March 31, 2004), only 141 of 7592 general practitioners in Quebec billed a total of 3209 spirometries performed on 2205 individuals  $\geq 55$  years of age (RAMQ, personal communication). Spirometry is, therefore, most often performed in hospital-based pulmonary function laboratories, ie, in the same hospitals that we surveyed and from which we obtained the data.

Our final citation belongs to Professor Austin Bradford Hill (1897-1991), English epidemiologist and statistician, who pioneered the randomized clinical trial and, together with Richard Doll, was the first to demonstrate the connection between cigarette smoking and lung cancer (30,31): "One must go seek more facts, paying less attention to techniques of handling the data and far more to the development and perfection of methods of obtaining them" (32). Accordingly, we strongly support the routine ascertainment of the validity of any diagnoses before using administrative databases for clinical and health services research.

---

**ACKNOWLEDGEMENTS:** The authors thank the archivists who kindly contributed to this study.

---

## REFERENCES

1. Baron JA, Weiderpass E. An introduction to epidemiological research with medical databases. *Ann Epidemiol* 2000;10:200-4.
2. Hennessy S. Use of health care databases in pharmacoepidemiology. *Basic Clin Pharmacol Toxicol* 2006;98:311-3.
3. Schneeweiss S, Avorn J. A review of uses of health care utilization databases for epidemiologic research on therapeutics. *J Clin Epidemiol* 2005;58:323-37.
4. Mayo NE, Nadeau L, Daskalopoulou SS, Cote R. The evolution of stroke in Quebec: A 15-year perspective. *Neurology* 2007;68:1122-7.
5. Pilon D, Castilloux AM, Dorais M, LeLorier J. Oral anticoagulants and the risk of osteoporotic fractures among elderly. *Pharmacoepidemiol Drug Saf* 2004;13:289-94.
6. Ernst P, Gonzalez AV, Brassard P, Suissa S. Inhaled corticosteroid use in chronic obstructive pulmonary disease and the risk of hospitalization for pneumonia. *Am J Respir Crit Care Med* 2007;176:162-6.
7. Keyhan G, Chen SF, Pilote L. The effectiveness of beta-blockers in women with congestive heart failure. *J Gen Intern Med* 2007;22:955-61.
8. Rabe KF, Hurd S, Anzueto A, et al. Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease: GOLD executive summary. *Am J Respir Crit Care Med* 2007;176:532-55.
9. Cochran WG. *Sampling Techniques*, 3rd edn. New York: John Wiley & Sons, 1977.
10. Kramer MS, Feinstein AR. Clinical biostatistics. LIV. The biostatistics of concordance. *Clin Pharmacol Ther* 1981;29:111-23.
11. Calverley PM, Anderson JA, Celli B, et al. Salmeterol and fluticasone propionate and survival in chronic obstructive pulmonary disease. *N Engl J Med* 2007;356:775-89.
12. Aaron SD, Vandemheen KL, Fergusson D, et al. Tiotropium in combination with placebo, salmeterol, or fluticasone-salmeterol for treatment of chronic obstructive pulmonary disease: A randomized trial. *Ann Intern Med* 2007;146:545-55.
13. Wedzicha JA, Calverley PM, Seemungal TA, Hagan G, Ansari Z, Stockley RA. The prevention of chronic obstructive pulmonary disease exacerbations by salmeterol/fluticasone propionate or tiotropium bromide. *Am J Respir Crit Care Med* 2008;177:19-26.

14. Fishman A, Martinez F, Naunheim K, et al. A randomized trial comparing lung-volume-reduction surgery with medical therapy for severe emphysema. *N Engl J Med* 2003;348:2059-73.
  15. Tashkin DP, Celli B, Senn S, et al. A 4-year trial of tiotropium in chronic obstructive pulmonary disease. *N Engl J Med* 2008;359:1543-54.
  16. Maltais F, Bourbeau J, Shapiro S, et al. Effects of home-based pulmonary rehabilitation in patients with chronic obstructive pulmonary disease: A randomized trial. *Ann Intern Med* 2008;149:869-78.
  17. Celli BR, Halbert RJ, Nordyke RJ, Schau B. Airway obstruction in never smokers: Results from the Third National Health and Nutrition Examination Survey. *Am J Med* 2005;118:1364-72.
  18. Chilvers ER, Lomas DA. Diagnosing COPD in non-smokers: Splitting not lumping. *Thorax* 2010;65:465-6.
  19. Holleman DR Jr, Simel DL. Does the clinical examination predict airflow limitation? *JAMA* 1995;273:313-9.
  20. Mannino DM, Gagnon RC, Petty TL, Lydick E. 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.
  21. Buist AS, McBurnie MA, Vollmer WM, et al. International variation in the prevalence of COPD (the BOLD Study): A population-based prevalence study. *Lancet* 2007;370:741-50.
  22. Chapman KR, Tashkin DP, Pye DJ. Gender bias in the diagnosis of COPD. *Chest* 2001;119:1691-5.
  23. Han MK, Postma D, Mannino DM, et al. Gender and chronic obstructive pulmonary disease: Why it matters. *Am J Respir Crit Care Med* 2007;176:1179-84.
  24. Jones RC, Dickson-Spillmann M, Mather MJ, Marks D, Shackell BS. Accuracy of diagnostic registers and management of chronic obstructive pulmonary disease: The Devon primary care audit. *Respir Res* 2008;9:62.
  25. Campbell SE, Campbell MK, Grimshaw JM, Walker AE. A systematic review of discharge coding accuracy. *J Public Health Med* 2001;23:205-11.
  26. Wilchesky M, Tamblyn RM, Huang A. Validation of diagnostic codes within medical services claims. *J Clin Epidemiol* 2004;57:131-41.
  27. Rutten FH, Zuithoff NP, Hak E, Grobbee DE, Hoes AW. Beta-blockers may reduce mortality and risk of exacerbations in patients with chronic obstructive pulmonary disease. *Arch Intern Med* 2010;170:880-7.
  28. McKnight J, Scott A, Menzies D, Bourbeau J, Blais L, Lemiere C. A cohort study showed that health insurance databases were accurate to distinguish chronic obstructive pulmonary disease from asthma and classify disease severity. *J Clin Epidemiol* 2005;58:206-8.
  29. Lacasse Y, Montori VM, Maltais F. Administrative database: Validity of recording vs. validity of diagnosis. *J Clin Epidemiol* 2006;59:104-5.
  30. Doll R, Hill AB. Smoking and carcinoma of the lung; preliminary report. *Br Med J* 1950;2:739-48.
  31. Doll R, Hill AB. Mortality in relation to smoking: Ten years' observations of British doctors. *BMJ* 1964;1:1460-7.
  32. Hill AB. Observation and experiment. *N Engl J Med* 1953;248:995-1001.
-