Open Access Full Text Article

ORIGINAL RESEARCH

Validity of ICD9-CM codes to diagnose chronic obstructive pulmonary disease from National Health Insurance claim data in Taiwan

Te-Wei Ho[1,2]
Sheng-Yuan Ruan[3]
Chun-Ta Huang[3,4]
Yi-Ju Tsai[5]
Feipei Lai[1]
Chong-Jen Yu[3]

[1]Graduate Institute of Biomedical Electronics and Bioinformatics, College of Electrical Engineering and Computer Science, National Taiwan University, Taipei, Taiwan; [2]Department of Surgery, College of Medicine, National Taiwan University, Taipei, Taiwan; [3]Department of Internal Medicine, National Taiwan University Hospital, Taipei, Taiwan; [4]Graduate Institute of Clinical Medicine, National Taiwan University, Taipei, Taiwan; [5]Graduate Institute of Biomedical and Pharmaceutical Science, College of Medicine, Fu Jen Catholic University, New Taipei City, Taiwan

Correspondence: Chun-Ta Huang
Department of Internal Medicine, National Taiwan University Hospital, 7 Chung-Shan South Road, Taipei 100, Taiwan
Tel +886 22 356 2557
Fax +886 22 322 2890
Email huangct@ntu.edu.tw

Purpose: Claim data from Taiwan's National Health Insurance (NHI) database have previously been utilized in the study of COPD. However, there are limited data on the positive predictive value of claim data for COPD diagnosis. Therefore, this study aimed to characterize and validate the COPD cohort identified from the NHI research database.

Methods: This cross-sectional study compared records from claim data with those from a medical center. From 2007 to 2014, a COPD cohort was constructed from claim data using ICD9-CM codes for COPD. The diagnostic positive predictive value of these data was assessed with reference to physician-verified COPD. In addition, a multivariate logistic regression model was built to identify independent factors associated with the positive predictive value of COPD diagnosis by claim data.

Results: During the 8-year study period, a total of 12,127 subjects met the criterion of having two or more outpatient codes in 1 year or one or more inpatient COPD codes in their claim data. Of this total, the diagnosis of COPD was verified by physicians in 7,701 (63.5%) subjects. Applying a more stringent criterion – three or more outpatient codes or two or more inpatient codes – improved the diagnostic positive predictive value to 72.2%. Age ≥65 years and a claim for spirometry were the two most important factors associated with the positive predictive value of claim-data-defined COPD. Adding spirometry testing to diagnostic ICD9-CM codes for COPD increased the positive predictive value to 84.6%.

Conclusion: This study emphasizes the importance of validation of disease-specific diagnosis prior to applying an administrative database in clinical studies. It also indicates the limitation of ICD9-CM codes alone in recognizing COPD patients within the NHI research database.

Keywords: chronic obstructive pulmonary disease, database, International Classification of Diseases code, Taiwan, validity

Abbreviations

CCI, Charlson Comorbidity Index; NTUH-IMD, National Taiwan University Hospital Integrated Medical Database; NHI, National Health Insurance; NHIRD, NHI Research Database.

Introduction

Though characterized as a preventable and treatable disease, COPD remains a leading cause of morbidity and mortality worldwide. Its prevalence and burden on health care systems are projected to increase over the coming decades.[1] In Taiwan, COPD was the seventh leading cause of death in 2012, with annual direct medical expenditure of about NT$4 billion.[2] Given its economic impact and public health implications,

research into the epidemiology, pathophysiology, biology, pharmacology, genetics, and health service delivery of COPD is ongoing.[3] Outcomes of COPD studies may vary based on ethnic, geographic, and racial differences.[4] Therefore, besides pursuing a global common ground, data describing regional and national differences are of equal importance and required to inform specific health care organizations and providers.

In 1995, Taiwan launched a single-payer National Health Insurance (NHI) program. As of 2014, over 99% of Taiwan's population was enrolled.[5] The NHI Research Database (NHIRD) is a population-based database established by the National Health Research Institute. It is based on claim data from the NHI program. Insurance claim data have increasingly been adopted in a variety of research fields. A research database provides such advantages as minimal referral bias, nonresponse, dropouts, and long follow-up periods, along with lower data collection costs. In the past few years, the NHIRD has been used to study several issues regarding COPD.[6–10] Diagnostic codes, such as the ICD9-CM codes, are commonly used to identify COPD patients. However, there are few data to validate the accuracy of ICD9-CM codes for COPD diagnosis using this specific database. It is important that interpreting claim data for COPD studies is standardized to ensure scientific validity and applicability across studies. Furthermore, a study has suggested that the accuracy rate of disease-specific claim data may significantly increase if the data are modified by valid methods.[11]

Therefore, the objective of this study was to characterize and validate the COPD cohorts identified from the NHIRD by a variety of criteria for applying ICD9-CM codes. Furthermore, we sought to establish if additional information, such as age, sex, comorbidities, and spirometry, could improve the predictive power of diagnostic codes to identify patients with COPD from within a large administrative claim database.

Methods
Study design and data source

This retrospective, cross-sectional study was conducted at National Taiwan University Hospital (NTUH), a 2,600-bed tertiary-care referral center in Northern Taiwan. From January 2007 to December 2014, findings from both the NTUH Integrated Medical Database (NTUH-IMD) and reimbursement claim data from the Taiwan NHI program at the NTUH were used in this study. To supplement information gathered from the NTUH-IMD and for data accuracy and integrity, chart reviews were conducted as needed. The NTUH-IMD is maintained by the NTUH. It contains all records of inpatient and outpatient visits to the NTUH since

2006, including data not covered by NHI. The claim data obtained from the NTUH were virtually identical to those retrieved from the NHIRD. The Research Ethics Committee of the NTUH approved the study protocol and waived off the need for informed consent, because of the retrospective nature of the study and any identifying patient data being only accessible to and kept confidential by the investigators.

Study participants

During the study period, a COPD cohort was constructed from the claim data using a common criterion, ie, at least two outpatient claims within a year or at least one inpatient claim coded for COPD (ICD9-CM codes 491, 492, and 496) in the first three or five diagnostic codes, respectively.[6–9] No restrictions on age or other parameters were applied in this cohort, because this study focused on examining features of COPD patients identified by diagnostic codes alone in claim data. We defined the index date as the date a patient received his or her first eligible diagnostic code.

Data collection

Patient demographics, comorbidities, and spirometry test results were retrieved from the claim data, while smoking status and spirometry results were obtained from the NTUH-IMD or chart reviews as appropriate. Comorbidities of interest were selected based on their common association and interactions with COPD.[1] They included hypertension (401.x–405.x), diabetes mellitus (250.x), cerebrovascular disease (430.x–438.x), heart failure (428.x), and coronary artery disease (410.x–414.x). These conditions were identified based on their corresponding ICD9-CM codes. The Charlson Comorbidity Index (CCI) score was calculated after excluding COPD as a comorbidity according to Deyo et al's description.[12] Demographics and smoking status were determined on the index date, and comorbidities were established during the year following the index date. Spirometry test results were collected between the preceding 2 years to 6 months following the index date.

Verification of COPD and patient classification

As spirometry is infrequently used in real-world practice for diagnosis of COPD, this study used physician-verified COPD, not spirometry-confirmed COPD, as the gold standard for COPD diagnosis.[13–15] Two pulmonologists with extensive COPD management experience who were blinded to grouping status independently reviewed the available information using ICD9-CM codes to classify patients into

three categories: COPD, indeterminate, and not COPD. Clinical manifestations, smoking history, and spirometry data were all taken into consideration for patient classification. When both pulmonologists agreed on a patient's category, the patient was classified as such. Otherwise, a patient's categorization was determined after a comprehensive discussion between the two physicians. Patients were classified as "indeterminate" when they could not be categorized clearly as having or not having COPD. In general, an "indeterminate" categorization indicated a lack of pertinent information for differential diagnosis of COPD and other respiratory diseases. For instance, an elderly patient with typical clinical features of COPD, a positive history of smoking and asthma, but with no spirometry test would be classified as indeterminate. A middle-aged patient with characteristic presentations of COPD, a history of bronchiectasis, and spirometry-confirmed airflow limitation could also be classified as indeterminate, because of uncertainties about the level of exposure to risk factors for COPD. When spirometry results were available, COPD staging followed the 2011 GOLD guidelines.[1] Spirometry-confirmed COPD was also defined based on these guidelines.

Study end points

The primary end point of this study was to validate the positive predictive value of COPD diagnosis in patient cohorts based on COPD-related ICD9-CM codes from claim data at the NTUH. Several criteria (two or more, three or more, or four or more outpatient vs one or more or two or more inpatient codes) were applied to construct different COPD cohorts. Other end points included identifying patient features associated with the diagnostic positive predictive value of COPD and proposing practical improvement measures.

Statistical analysis

Continuous variables are reported as means ± SD and compared using Student's t-test or one-way ANOVA as appropriate. Categorical variables were summarized by percentages and analyzed by χ^2 test. To identify independent factors associated with the positive predictive value of COPD diagnosis by claim data, a multivariate logistic regression model was constructed and ORs reported with 95% CIs. Dichotomized variables with a P-value <0.05 in the univariate analysis were entered into the multivariate model. In this study, smoking status was not entered into the final model, because it could not be properly identified from the claim data.[16]

Interobserver agreement on COPD diagnosis between the two pulmonologists was quantified by κ-coefficient,

where κ-values of 0.01–0.20 indicated slight agreement, 0.21–0.40 fair agreement, 0.41–0.60 moderate agreement, 0.61–0.80 substantial agreement, and 0.81–0.99 almost perfect agreement.[17] Microsoft SQL Server 2008 was used for data management and computing. Data analysis was performed using SPSS version 22.0 (IBM, Armonk, NY, USA). A two-tailed P-value <0.05 indicated statistical significance. Figures were plotted with Microsoft Office Excel 2013.

Results

Study population

During the 8-year study period, 12,127 subjects fulfilled the criterion of having two or more outpatient codes within any single year or one or more inpatient COPD codes in their claim data (Figure 1). The average age of this COPD cohort was 70.0±12.2 years and approximately three quarters were male. Smokers or ex-smokers comprised 70.9% of the study cohort. Hypertension (33.4%) and coronary artery disease (17.7%) were the two most common comorbidities. Interobserver agreement on COPD diagnosis between the two pulmonologists was substantial: κ=0.77 (95% CI 0.75–0.79, P<0.001) with a 92% concordance rate. Using data from the NTUH-IMD and chart reviews, COPD diagnosis was verified by the pulmonologists in 63.5% of the cohort, while 24.1% of the cohort was considered not to have COPD. Spirometry-confirmed COPD was observed in 49.2% of the COPD cohort. The proportions of physician-verified
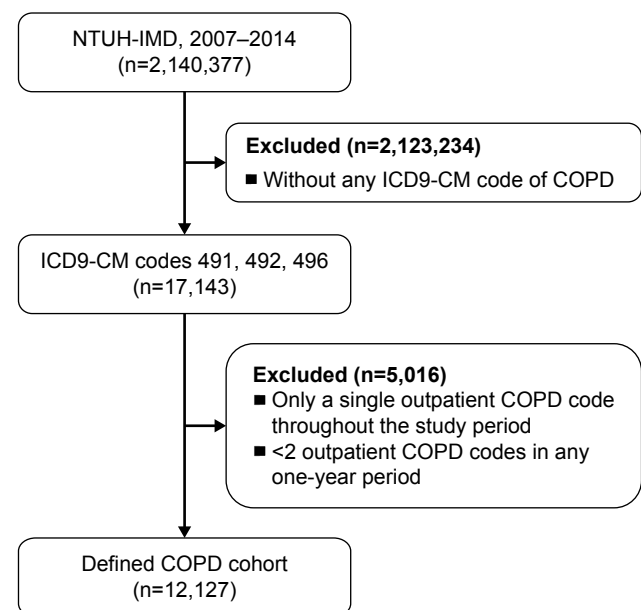

Figure 1 Study flow diagram.
Abbreviations: NTUH, National Taiwan University Hospital; IMD, Integrated Medical Database.

COPD patients in the study cohort according to age, sex, and smoking status are displayed in Table S1. In addition, Table S2 compares clinical features between spirometry-confirmed COPD and other patients in the cohort.

Features of COPD cohorts defined according to claim data

Table 1 shows the characteristics of COPD cohorts identified from claim data, with a variety of criteria for COPD-related ICD9-CM codes. Although there were several differences in demographics and comorbidities across the patient cohorts, the stricter criterion generated a COPD cohort with a higher proportion (63.5%–72.2%) of verified COPD patients. In addition, spirometry-confirmed COPD patients were more commonly seen (49.2%–57.0%) in this cohort. However, the distribution of GOLD stages 1–4 did not differ significantly among the COPD cohorts.

Variables associated with diagnostic accuracy of COPD by claim data

In the cohort of patients with three or more outpatient codes within 1 year or two or more inpatient COPD diagnostic codes in their claim data (Table 2), we observed that physician-verified COPD patients were more likely to be older (71.4 vs 67.1 years, $P<0.001$), male (77.9% vs 66.5%, $P<0.001$), and have a higher CCI score (0.98 vs 0.77, $P<0.001$) than other patients. Moreover, a higher proportion of physician-verified COPD patients were smokers or ex-smokers (77.4% vs 58.1%, $P<0.001$) and had undergone spirometry testing (81.1% vs 46.4%, $P<0.001$) compared to other patients. In the multivariate analysis (Table 3), age ≥65 years (OR 2.363), male sex (OR 1.410), CCI score (excluding COPD) ≥1 (OR 1.301), heart failure (OR 1.330), and having had spirometry performed (OR 5.656) were independent factors associated with accurate COPD

Table 1 Comparisons of COPD patient cohorts identified from claim data using different criteria of ICD9-CM codes

	ICD9-CM COPD codes					*P*-value
	≥2 outpatients or ≥1 inpatients	≥2 outpatients or ≥2 inpatients	≥3 outpatients or ≥1 inpatients	≥4 outpatients or ≥1 inpatients	≥3 outpatients or ≥2 inpatients	
	n=12,127	n=10,699	n=10,053	n=8,837	n=8,485	
Age, years	70.0±12.2	69.4±12.1	70.9±11.9	71.5±11.7	70.2±11.9	<0.001
≥65	7,996 (65.9)	6,857 (64.1)	6,959 (69.2)	6,310 (71.4)	5,708 (67.3)	<0.001
Male sex	8,971 (74.0)	7,855 (73.4)	7,574 (75.3)	6,753 (76.4)	6,343 (74.8)	<0.001
Smoking status						
Smoker/ex-smoker	8,604 (70.9)	7,295 (68.2)	7,556 (75.2)	6,888 (77.9)	6,114 (72.1)	<0.001
Never smoker	1,762 (14.5)	1,747 (16.3)	1,292 (12.9)	1,032 (11.7)	1,276 (15.0)	
Unknown	1,761 (14.5)	1,657 (15.5)	1,205 (12.0)	917 (10.4)	1,095 (12.9)	
CCI score, excluding COPD	0.91±1.52	0.86±1.47	0.97±1.56	1.01±1.58	0.92±1.50	<0.001
0	7,051 (58.1)	6,346 (59.3)	5,599 (55.7)	4,776 (54.0)	4,828 (56.9)	<0.001
≥1	5,076 (41.9)	4,353 (40.7)	4,454 (44.3)	4,061 (46.0)	3,657 (43.1)	
Comorbidities						
Hypertension	4,053 (33.4)	3,691 (34.5)	3,493 (34.7)	3,119 (35.3)	3,078 (36.3)	0.001
Coronary artery disease	2,145 (17.7)	1,894 (17.7)	1,850 (18.4)	1,666 (18.9)	1,567 (18.5)	0.130
Diabetes mellitus	1,736 (14.3)	1,524 (14.2)	1,509 (15.0)	1,367 (15.5)	1,271 (15.0)	0.074
Heart failure	1,238 (10.2)	1,018 (9.5)	1,094 (10.9)	1,000 (11.3)	845 (10.0)	<0.001
Cerebrovascular disease	1,120 (9.2)	943 (8.8)	994 (9.9)	912 (10.3)	794 (9.4)	0.003
Physician verification						
COPD	7,701 (63.5)	7,006 (65.5)	6,923 (68.9)	6,330 (71.6)	6,125 (72.2)	<0.001
Indeterminate	1,502 (12.4)	987 (9.2)	1,261 (12.5)	1,136 (12.9)	729 (8.6)	
Not COPD	2,924 (24.1)	2,706 (25.3)	1,869 (18.6)	1,371 (15.5)	1,631 (19.2)	
Spirometry-confirmed COPD	5,963 (49.2)	5,531 (51.7)	5,340 (53.1)	4,874 (55.2)	4,838 (57.0)	<0.001
Spirometry done	n=7,914	n=7,341	n=6,593	n=5,781	n=5,933	
GOLD stage						
I	2,682 (45.0)	2,532 (45.8)	2,309 (43.2)	2,049 (42.0)	2,130 (44.0)	0.062
II	2,279 (38.2)	2,084 (37.7)	2,084 (39.0)	1,930 (39.6)	1,864 (38.5)	
III	844 (14.2)	775 (14.0)	793 (14.9)	746 (15.3)	710 (14.7)	
IV	158 (2.6)	140 (2.5)	154 (2.9)	149 (3.1)	134 (2.8)	

Abbreviation: CCI, Charlson Comorbidity Index.

Table 2 Characteristics of patient cohorts with three or more outpatient or two or more inpatient ICD9-CM COPD codes

	COPD	Indeterminate/ not COPD	P-value
	n=6,125	n=2,360	
Age, years	71.4±11.4	67.1±12.5	<0.001
≥65	4,400 (71.8)	1,308 (55.4)	<0.001
Male gender	4,774 (77.9)	1,569 (66.5)	<0.001
Smoking status			
Smoker/ex-smoker	4,742 (77.4)	1,372 (58.1)	<0.001
Never smoker	734 (12.0)	542 (23.0)	
Unknown	649 (10.6)	446 (18.9)	
CCI score, excluding COPD	0.98±1.55	0.77±1.33	<0.001
0	3,386 (55.3)	1,442 (61.1)	<0.001
≥1	2,739 (44.7)	918 (38.9)	
Comorbidities			
Hypertension	2,197 (35.9)	881 (37.3)	0.210
Coronary artery disease	1,152 (18.8)	415 (17.6)	0.193
Diabetes mellitus	904 (14.8)	367 (15.6)	0.360
Heart failure	665 (10.9)	180 (7.6)	<0.001
Cerebrovascular disease	589 (9.6)	205 (8.7)	0.188
Spirometry performed	4,970 (81.1)	1,095 (46.4)	<0.001

Abbreviation: CCI, Charlson Comorbidity Index.

diagnosis by claim data when evaluated against the GOLD standard – physician diagnosis.

Patient cohorts were further divided into four subgroups using the two strongest diagnostic accuracy-associated factors: age and spirometry (Figure 2). If spirometry had not been performed, COPD diagnosis was physician-verified in less than half the patients in each cohort. Instead, >80% of COPD patients were physician-verified in the subgroup of patients aged ≥65 years and with spirometry performed across different cohorts. Therefore, including spirometry testing as a prerequisite for diagnostic codes for COPD in claim data identified a significantly higher proportion of physician-verified COPD patients (Figure 3).

Discussion

This study found that the diagnostic positive predictive value defined by ICD9-CM codes alone using a common criterion

Table 3 Multivariate logistic regression analysis of clinical features abstracted from claim data with the physician-verified COPD diagnosis as the outcome variable

	OR (95% CI)	P-value
Age ≥65 years	2.363 (2.117–2.638)	<0.001
Male sex	1.410 (1.257–1.582)	<0.001
CCI score (excluding COPD) ≥1	1.301 (1.159–1.460)	<0.001
Heart failure	1.330 (1.089–1.625)	0.005
Spirometry performed	5.656 (5.065–6.315)	<0.001

Abbreviation: CCI, Charlson Comorbidity Index.

of two or more outpatient codes in 1 year or one or more inpatient codes on patient data from the NHIRD in Taiwan was 63.5%. Applying a more stringent criterion – three or more outpatient codes in 1 year or two or more inpatient codes – improved the diagnostic positive predictive value to 72.2%. Performance of spirometry and age ≥65 years were the two most important factors associated with the positive predictive value of claim data-defined COPD. Adding spirometry testing as a prerequisite to diagnostic codes for COPD in claim data evidently increased the positive predictive value to 84.6% (Figure 3) compared to ICD9-CM codes alone (63.5%–72.2%). Therefore, this study suggests that it may not be appropriate to define COPD patients in the NHIRD using diagnostic ICD9-CM codes alone, given their relatively low diagnostic positive predictive value. The addition of spirometry to preexisting ICD9-CM codes and criteria can significantly boost the accuracy of COPD diagnosis.

The level of clinical detail encoded by the ICD9-CM coding system, problems related to coding accuracy, and financial incentives may all limit the suitability of insurance claim databases for use in research.[18] As such, routine assessment of the validity of disease-specific diagnosis prior to using an administrative database in clinical and health care service research has been advocated.[19] To our knowledge, this study is the first to validate the positive predictive value of COPD diagnosis using the NHIRD in Taiwan. The positive predictive value from this study was within the range (50.4%–92.0%) reported in the literature.[20–22] Differences in the target population validated (inpatient vs outpatient vs inpatient plus outpatient) the gold standard applied (physician diagnosis vs spirometry-verified), and the primary reason for building the database (billing purposes vs managerial decision making) may account for the wide range of diagnostic positive predictive values across several studies. Despite these variations, our study is valuable, because it paves the way for COPD studies to be conducted using the NHIRD in Taiwan.

This study was performed in a single medical center. Therefore, concerns may exist about the generalizability of findings from this study. However, our results may be applicable to other medical centers in Taiwan, as they were accredited with the NHI program based on the same rules, policies, and procedures. Secondly, the data-processing system for reimbursement is identical across all health care providers of the same accreditation level. Furthermore, all NHI-contracted health care facilities are under regular inspection by the NHI administration to prevent illegal practices
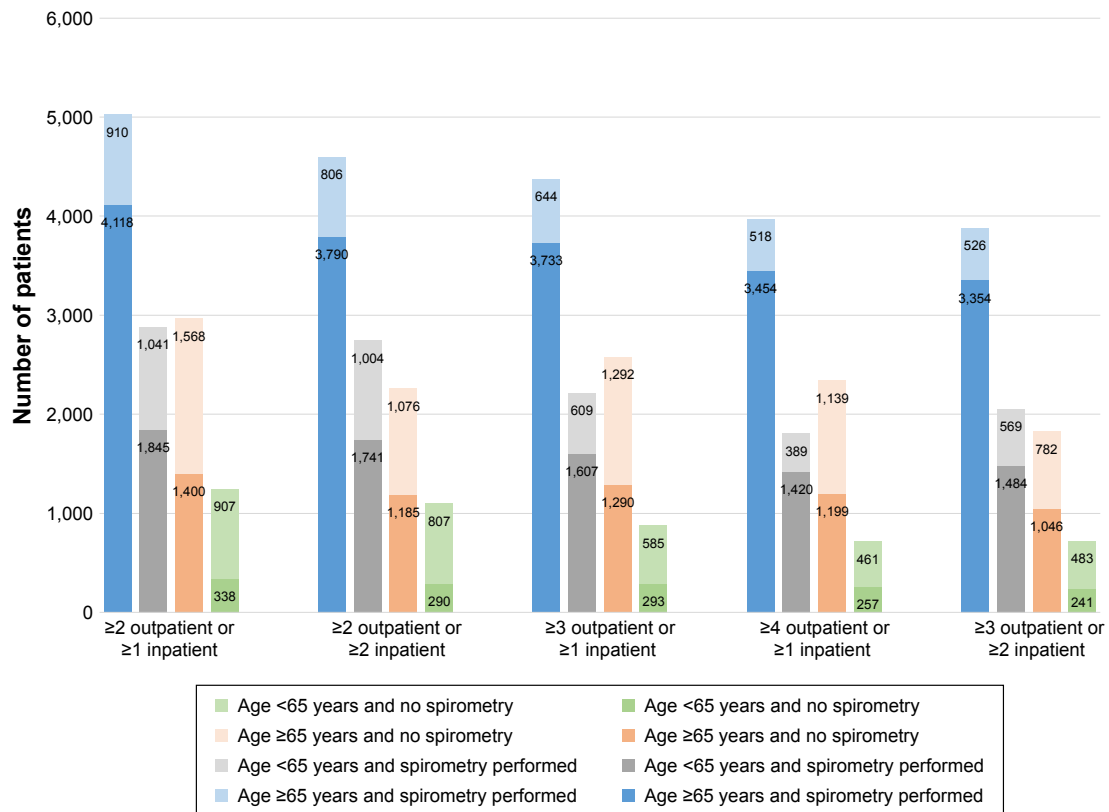
Figure 2 Age and spirometry-related data for verified COPD patients in each cohort defined by claim data.
Note: For each bar, the darker color indicates the proportion of COPD patients verified by physicians with reference to the specified cohort defined by COPD codes, age, and spirometry.
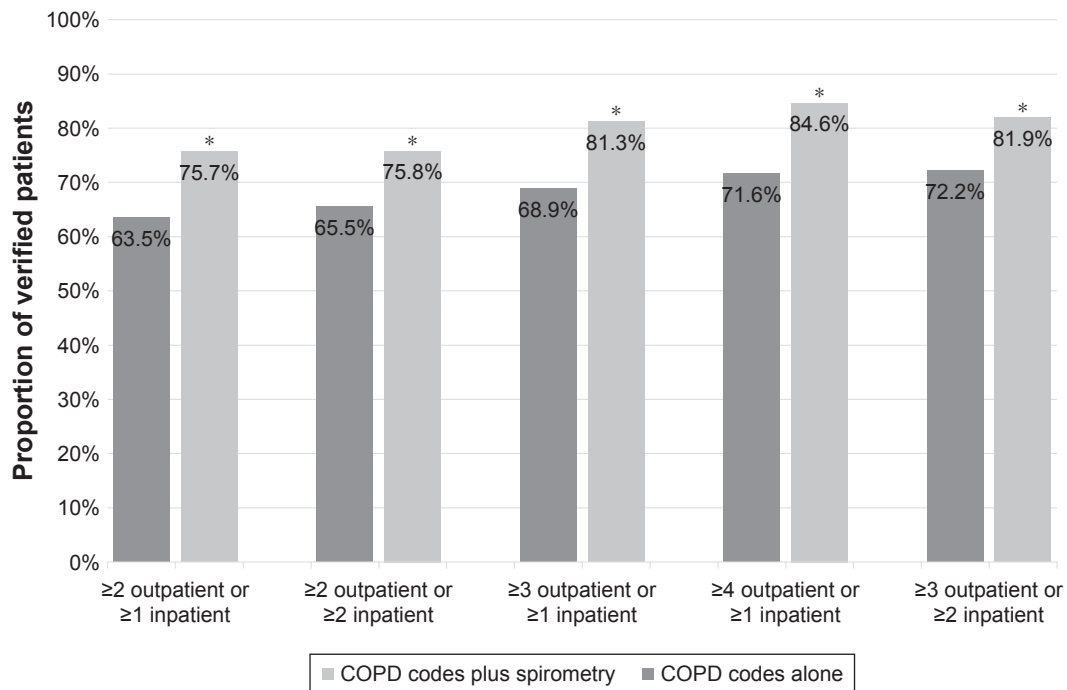

Figure 3 Comparisons of proportion of physician-verified COPD patients in each cohort defined by COPD codes alone or COPD codes plus spirometry in claim data.
Note: *$P<0.001$ as compared to COPD codes alone.

or unsatisfactory quality.[23] Several studies conducted in Taiwan have suggested that the virtually barrier-free health care system provided by the NHI program significantly diminishes financial barriers to care,[24] improves resource allocation,[25] and reduces geographic disparity in health status or utilization.[26] All these factors ensure comparable quality and reliability of care across different levels of service providers, and may extend the applicability of our findings to lower-level facilities. In addition, similarities in the distribution of demographics and comorbidities between our cohort and other COPD cohorts in epidemiological and population-based studies in Taiwan support the generalizability of the present study.[6,8,9,27] However, a study in Taiwan found that coding quality is generally better in medical centers and regional hospitals compared to district hospitals.[28] Therefore, the diagnostic validity of COPD in our study might be higher than the average. As a result, findings from this study still need to be applied and interpreted with some caution.

Spirometry is the best standardized and most objective measurement of airflow limitation available and is crucial in COPD diagnosis.[1] However, it has been shown that only a third of newly diagnosed COPD patients undergo spirometry to confirm the presence of airflow obstruction.[13–15] Accordingly, most studies utilize chart reviews or physician verification as the gold standard to verify the validity of ICD9-CM codes for COPD.[20,29–31] In this study, approximately half the COPD cohorts had spirometry performed, a figure close to that obtained from recent studies in the US and Sweden.[32,33] Advanced age has been identified as a factor in the underuse of spirometry. This might explain the underutilization of this modality in an elderly population like ours.[14,15] Several other factors, such as physician attitude and practice patterns, time constraints during clinic visits, and confidence in clinical judgments,[34–36] have been identified as reasons for the underutilization of spirometry in clinical practice. However, these subjective issues are difficult to evaluate through chart reviews. In our study, a claim for spirometry was recognized as the single most important aid to ICD9-CM codes in the correct identification of COPD cases from the NHIRD. Therefore, it should be incorporated into the common criteria in future studies of this kind. This also supports the pivotal position of spirometry in clinical diagnosis of COPD.

In this study, age was another significant factor associated with the diagnostic positive predictive value of COPD in claim data using ICD9-CM codes. This may be because the prevalence of COPD increases with age.[37,38] Furthermore, elderly patients usually have multiple comorbidities and are more likely to have their COPD diagnosed, because of increased physician encounters and diagnostic procedures. Therefore, our findings suggest that it is more reliable to study COPD in the aged population using the NHIRD.

Limitations to this study must be considered. Administrative claim data of the NHIRD are primarily generated for imbursement rather than for research purposes. Therefore, the content and quality of the data may change over time in response to regulations by the NHI administration. Validation studies for a given disease should be conducted regularly if a database is to be used repeatedly over different time periods. While our study spans a long time frame, 2007–2014, caution needs to be exercised when extrapolating our results before and beyond the study years.

Conclusion

In summary, our study validated the positive predictive value of COPD claim data from the NHIRD for the first time. It also suggested that it may not be satisfactory for researchers to identify COPD subjects using only ICD9-CM codes. Age ≥65 years and performance of spirometry were significant factors associated with accurate COPD diagnosis by claim data. Adding spirometry testing as a prerequisite for diagnostic ICD9-CM codes for COPD evidently improved the diagnostic positive predictive value. The findings of this study emphasize the importance of validation of disease-specific diagnosis prior to applying an administrative database in clinical studies, and indicate the limitation of ICD9-CM codes alone in recognizing COPD patients in the NHIRD.

Acknowledgments

Disclosure

The authors report no conflicts of interest in this work.

References

1. Vestbo J, Hurd SS, Agustí AG, et al. Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease: GOLD executive summary. *Am J Respir Crit Care Med*. 2013;187(4): 347–365.

2. Ministry of Health and Welfare, Executive Yuan, R.O.C. (TAIWAN). Available from: http://www.mohw.gov.tw/cp-3218-22790-1.html. Accessed January 24, 2018.

3. Maclay JD, Rabinovich RA, Macnee W. Update in chronic obstructive pulmonary disease 2008. *Am J Respir Crit Care Med*. 2009;179(7): 533–541.

4. Gilkes A, Ashworth M, Schofield P, et al. Does COPD risk vary by ethnicity? A retrospective cross-sectional study. *Int J Chron Obstruct Pulmon Dis*. 2016;11:739–746.

5. National Health Insurance Administration, Ministry of Health and Welfare, Taiwan, R.O.C. (2014). *National Health Insurance Annual Report*. 2014–2015.

6. Ho TW, Huang CT, Ruan SY, Tsai YJ, Lai F, Yu CJ. Diabetes mellitus in patients with chronic obstructive pulmonary disease-The impact on mortality. *PLoS One*. 2017;12(4):e0175794.

7. Lee CT-C, Mao I-C, Lin C-H, Lin S-H, Hsieh M-C. Chronic obstructive pulmonary disease: a risk factor for type 2 diabetes: a nationwide population-based study. *Eur J Clin Invest*. 2013;43(11):1113–1119.

8. Wang MT, Lo YW, Tsai CL, et al. Statin use and risk of COPD exacerbation requiring hospitalization. *Am J Med*. 2013;126(7):e592:598–606.

9. Yang YW, Chen YH, Wang KH, Wang CY, Lin HW. Risk of herpes zoster among patients with chronic obstructive pulmonary disease: a population-based study. *CMAJ*. 2011;183(5):E275–E280.

10. Ho TW, Tsai YJ, Ruan SY, et al. In-hospital and one-year mortality and their predictors in patients hospitalized for first-ever chronic obstructive pulmonary disease exacerbations: a nationwide population-based study. *PLoS One*. 2014;9(12):e114866.

11. Quam L, Ellis LB, Venus P, Clouse J, Taylor CG, Leatherman S. Using claims data for epidemiologic research. The concordance of claims-based criteria with the medical record and patient survey for identifying a hypertensive population. *Med Care*. 1993;31(6):498–507.

12. Deyo RA, Cherkin DC, Ciol MA. Adapting a clinical comorbidity index for use with ICD-9-CM administrative databases. *J Clin Epidemiol*. 1992;45(6):613–619.

13. Joo MJ, Lee TA, Weiss KB. Geographic variation of spirometry use in newly diagnosed COPD. *Chest*. 2008;134(1):38–45.

14. Han MK, Kim MG, Mardon R, et al. Spirometry utilization for COPD: how do we measure up? *Chest*. 2007;132(2):403–409.

15. Lee TA, Bartle B, Weiss KB. Spirometry use in clinical practice following diagnosis of COPD. *Chest*. 2006;129(6):1509–1515.

16. Brookhart MA, Stürmer T, Glynn RJ, Rassen J, Schneeweiss S. Confounding control in healthcare database research: challenges and potential approaches. *Med Care*. 2010;48(6 Suppl):S114–S120.

17. Viera AJ, Garrett JM. Understanding interobserver agreement: the kappa statistic. *Fam Med*. 2005;37(5):360–363.

18. Mullin RL. Diagnosis-related groups and severity. ICD-9-CM, the real problem. *JAMA*. 1985;254(9):1208–1210.

19. Lacasse Y, Montori VM, Maltais F. Administrative database: validity of recording vs validity of diagnosis. *J Clin Epidemiol*. 2006;59(1): 104–105.

20. Thomsen RW, Lange P, Hellquist B, et al. Validity and underrecording of diagnosis of COPD in the Danish National Patient Registry. *Respir Med*. 2011;105(7):1063–1068.

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

22. Cooke CR, Joo MJ, Anderson SM, et al. The validity of using ICD-9 codes and pharmacy records to identify patients with chronic obstructive pulmonary disease. *BMC Health Serv Res*. 2011;11:37.

23. Ho Chan WS. Taiwan's healthcare report 2010. *EPMA J*. 2010;1(4): 563–585.

24. Cheng SH, Chiang TL. The effect of universal health insurance on health care utilization in Taiwan. Results from a natural experiment. *JAMA*. 1997;278(2):89–93.

25. Kreng VB, Yang CT. The equality of resource allocation in health care under the National Health Insurance System in Taiwan. *Health Policy*. 2011;100(2–3):203–210.

26. Huang N, Yip W, Chang HJ, Chou YJ. Trends in rural and urban differentials in incidence rates for ruptured appendicitis under the National Health Insurance in Taiwan. *Public Health*. 2006;120(11): 1055–1063.

27. Cheng SL, Chan MC, Wang CC, et al. COPD in Taiwan: a National Epidemiology Survey. *Int J Chron Obstruct Pulmon Dis*. 2015;10: 2459–2467.

28. Lai HT, Yaung CL, Fan BY. Study on the quality and its related factors of international classification of disease coding system in National Health Insurance. *Chin J Pub Health*. 1998;17:11.

29. Lacasse Y, Montori VM, Lanthier C, Maltis F. The validity of diagnosing chronic obstructive pulmonary disease from a large administrative database. *Can Respir J*. 2005;12(5):251–256.

30. Rawson NS, Malcolm E. Validity of the recording of ischaemic heart disease and chronic obstructive pulmonary disease in the Saskatchewan health care datafiles. *Stat Med*. 1995;14(24):2627–2643.

31. Mapel DW, Frost FJ, Hurley JS, et al. An algorithm for the identification of undiagnosed COPD cases using administrative claims data. *J Manag Care Pharm*. 2006;12(6):458–465.

32. Nishi SP, Wang Y, Kuo YF, Goodwin JS, Sharma G. Spirometry use among older adults with chronic obstructive pulmonary disease: 1999–2008. *Ann Am Thorac Soc*. 2013;10(6):565–573.

33. Arne M, Lisspers K, Ställberg B, et al. How often is diagnosis of COPD confirmed with spirometry? *Respir Med*. 2010;104(4):550–556.

34. Joo MJ, Sharp LK, Au DH, Lee TA, Fitzgibbon ML. Use of spirometry in the diagnosis of COPD: a qualitative study in primary care. *COPD*. 2013;10(4):444–449.

35. Walters JA, Hansen E, Mudge P, Johns DP, Walters EH, Wood-Baker R. Barriers to the use of spirometry in general practice. *Aust Fam Physician*. 2005;34(3):201–203.

36. Moore PL. Practice management and chronic obstructive pulmonary disease in primary care. *Am J Med*. 2007;120(8 Suppl 1):S23–S27.

37. Xu F, Yin X, Zhang M, Shen H, Lu L, Xu Y. Prevalence of physician-diagnosed COPD and its association with smoking among urban and rural residents in regional mainland China. *Chest*. 2005;128(4): 2818–2823.

38. Soriano JB, Maier WC, Egger P, et al. Recent trends in physician diagnosed COPD in women and men in the UK. *Thorax*. 2000;55(9): 789–794.

Supplementary materials

Table S1 Proportions of physician-verified COPD patients stratified by age, sex, and smoking status in the study cohort (n=12,127)

	COPD	Indeterminate	Not COPD	*P*-value
	n=7,701	n=1,502	n=2,924	
Age, years				
<65	2,183 (53)	469 (11)	1,479 (36)	<0.001
≥65	5,518 (69)	1,033 (13)	1,445 (18)	
Sex				
Male	5,884 (66)	1,137 (13)	1,950 (22)	<0.001
Female	1,817 (58)	365 (12)	974 (31)	
Smoking status				
Smoker/ex-smoker	5,998 (70)	1,462 (17)	1,144 (13)	<0.001
Never-smoker	883 (50)	18 (1.0)	861 (49)	
Unknown	820 (47)	22 (1.2)	919 (52)	

Table S2 Comparisons of clinical characteristics of spirometry-confirmed COPD and other patient cohorts (n=12,127)

	Spirometry-confirmed	Other cohorts	*P*-value
	n=5,963	n=6,164	
Age, years	70.4±11.2	69.7±13.2	0.003
≥65	4,118 (69.1)	3,878 (62.9)	<0.001
Male sex	4,714 (79.1)	4,257 (69.1)	<0.001
Smoking status			
Smoker/ex-smoker	4,621 (77.5)	3,983 (64.6)	<0.001
Never smoker	696 (11.7)	1,066 (17.3)	
Unknown	646 (10.8)	1,115 (18.1)	
CCI, excluding COPD	0.94±1.58	0.87±1.48	0.012
0	3,430 (57.5)	3,621 (58.7)	0.172
≥1	2,533 (42.5)	2,543 (41.3)	
Comorbidities			
Hypertension	1,998 (33.5)	2,055 (33.3)	0.845
Coronary artery disease	1,111 (18.6)	1,034 (16.8)	0.007
Diabetes mellitus	827 (13.9)	909 (14.7)	0.168
Heart failure	619 (10.4)	619 (10.0)	0.538
Cerebrovascular disease	445 (7.5)	657 (11.0)	<0.001

Abbreviation: CCI, Charlson Comorbidity Index.