Validation of claims-based algorithms for pulmonary arterial hypertension

Ravikanth Papani[1] [ORCID], Gulshan Sharma[1], Amitesh Agarwal[2], Sean J. Callahan[3], Winston J. Chan[4], Yong-Fang Kuo[4], Yun M. Shim[3], Andrew D. Mihalek[3] and Alexander G. Duarte[1]

[1]Division of Pulmonary, Critical Care, and Sleep Medicine, University of Texas Medical Branch, Galveston, TX, USA; [2]Division of Pulmonary, Critical Care, and Sleep Medicine, University of Florida College of Medicine, Jacksonville, FL, USA; [3]Division of Pulmonary and Critical Care Medicine, University of Virginia School of Medicine, Charlottesville, VA, USA; [4]Office of Biostatistics, Department of Preventive Medicine and Community Health, University of Texas Medical Branch, Galveston, TX, USA

Abstract

Administrative claims studies do not adequately distinguish pulmonary arterial hypertension (PAH) from other forms of pulmonary hypertension (PH). Our aim is to develop and validate a set of algorithms using International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) codes and electronic medical records (EMR), to identify patients with PAH. From January 2012 to August 2015, the EMRs of patients with ICD-9-CM codes for PH with an outpatient visit at the University of Texas Medical Branch were reviewed. Patients were divided into PAH or non-PAH groups according to EMR encounter diagnosis. Patient demographics, echocardiography, right heart catheterization (RHC) results, and PAH-specific therapies were assessed. RHC measurements were reviewed to categorize cases as hemodynamically determined PAH or not PAH. Weighted sensitivity, specificity, and positive and negative predictive values were calculated for the developed algorithms. A logistic regression analysis was conducted to determine how well the algorithms performed. External validation was performed at the University of Virginia Health System. The cohort for the development algorithms consisted of 683 patients with PH, PAH group (n = 191) and non-PAH group (n = 492). A hemodynamic diagnosis of PAH determined by RHC was recorded in the PAH (26%) and non-PAH (3%) groups. The positive predictive value for the algorithm that included ICD-9-CM and PAH-specific medications was 66.9% and sensitivity was 28.2% with a c-statistic of 0.66. The positive predictive value for the EMR-based algorithm that included ICD-9-CM, EMR encounter diagnosis, echocardiography, RHC, and PAH-specific medication was 69.4% and a c-statistic of 0.87. A validation cohort of 177 patients with PH examined from August 2015 to August 2016 using EMR-based algorithms yielded a similar positive predictive value of 62.5%. In conclusion, claims-based algorithms that included ICD-9-CM codes, EMR encounter diagnosis, echocardiography, RHC, and PAH-specific medications better-identified patients with PAH than ICD-9-CM codes alone.

Pulmonary hypertension (PH) is a complex condition leading to a progressive decline in daily functional activities and premature death. A clinical classification scheme was developed with subsequent modifications that led to the current five group classification that is based on etiology.[1] Importantly, this classification system has provided a structure whereby patients have been enrolled in clinical trials that have led to novel medical therapies[2,3] and registries[4–6] that have led to insights to the natural history of PH.

However, enrollment and participation in pharmaceutical trials and registries often reflect patient care in the context of the clinical trial and at referral centers. Moreover,

Corresponding author:
Ravikanth Papani, Division of Pulmonary, Critical Care, and Sleep Medicine, University of Texas Medical Branch, John Sealy Annex 5.140, 301 University Boulevard, Galveston, TX 77555, USA.
Email: Ravikanth.papani@gmail.com

management of pulmonary arterial hypertension (PAH) in the community setting has been reported to vary when compared to tertiary referral centers.[7,8] Thus, outside of clinical trials or patient registries there is a need to understand and assess health care use of patients with PAH.

The use of large administrative databases and electronic medical records (EMR) provides an opportunity to conduct population-based studies in PAH. Administrative claims data are generated at every encounter within a healthcare system that allows investigation into epidemiology, procedure utilization, drug prescription, hospitalization rate, and mortality associated with PAH.[9–13] Administrative claims databases rely on International Classification of Diseases Ninth Revision Clinical Modification (ICD-9-CM) and ICD -10 Revision codes that segregate patients into primary and secondary causes of PH, but this does not adequately reflect the clinical classification scheme.[14] While investigators have attempted to identify patients with PAH using ICD-9 and Current Procedural Terminology (CPT) codes, the ability of algorithms to discriminate patients with PAH from those with other forms of PH has not been adequately assessed. In this study, we used ICD-9-CM codes and other EMR data to develop algorithms and then assessed the discriminatory characteristics of the algorithm to identify patients with PAH.

Methods

We performed a retrospective study of patients with PH seen in outpatient clinics at the University of Texas Medical Branch and the University of Virginia. The EPIC® EMR system was used to collect patient information at both study sites. This observational study was classified as exempt research by the University of Texas Medical Branch and University of Virginia Institutional Review Board (IRB).

Algorithm development

From January 2012 to August 2015 (pre-ICD-10 era), we identified patients aged ≥ 18 years with outpatient claims that included ICD-9-CM codes for primary PH (416.0) or secondary PH (416.8) seen at the University of Texas Medical Branch. Patients with ICD-9-CM codes for PH were identified and, using the EPIC® EMR system, we collected demographics, echocardiography, and right heart catheterization (RHC) measurements and prescribed PAH-specific medications. Subsequently, patients were divided into two groups based on the initial EMR encounter diagnosis as entered by the physician at the time of visit, that was suggestive of PAH (PAH group) or not suggestive of PAH (non-PAH group) (Fig. 1). Ten phrases were identified in the EMR that represented idiopathic or conditions associated with PAH (Table 1) and 13 phrases that represented PH not due to PAH. For example, the phrases selected for the PAH group included the terms "primary pulmonary hypertension" or contained the term "pulmonary arterial hypertension." In contrast, the terms used to identify the non-PAH group did not contain the phrase "primary pulmonary hypertension" or "pulmonary arterial hypertension," rather they included "secondary pulmonary hypertension" or a severity of PH. In order to be all inclusive, we used broad terminologies to capture different ways providers enter a diagnosis of PAH or PH in the EMR. Although, there were ten different phrases suggestive of PAH, principally two terms were used. For example, the term "primary pulmonary hypertension" was used in three phrases while the remaining seven phrases used "pulmonary artery hypertension." A structured review of the EMRs was performed for all patients in the PAH group and for a 20% random sample in the non-PAH group. The following variables were collected and reviewed: (1) age, sex, race, body mass index, and co-morbidities; (2) echocardiography results; (3) RHC results, including mean pulmonary artery pressure (mPAP), pulmonary capillary wedge pressure, cardiac output, pulmonary vascular resistance (PVR), and vasodilator responsiveness; and (4) prescribed PAH-specific therapies categorized as phosphodiesterase-5 inhibitors, endothelin receptor antagonists, and prostanoids. Prescription of PAH-specific therapies did not automatically confer a diagnosis of PAH. Patients were diagnosed with PAH after review and interpretation of RHC measurements based on previously published guidelines.[1] A hemodynamic diagnosis of PAH was determined with RHC mPAP ≥ 25 mmHg and pulmonary capillary wedge pressure ≤ 15 mmHg. A PAH expert (AD) reviewed the clinical charts and patients with airflow limitation defined as FEV1/FVC < 0.70 were excluded. Patients with restrictive lung disease defined as FVC $< 60\%$ predicted and PaO2 < 60 mmHg were also excluded. A structured chart review was performed to assess for CTEPH and a review of CTA and/or VQ scan results was performed whenever these imaging studies available. Patients were given a score of 0 if the scan was negative and a score of 1 if positive; patients with positive results were excluded. Subsequently, algorithms were developed using ICD-9-CM codes, EMR encounter diagnosis, presence of echocardiography, presence of RHC, and a prescription for PAH-specific therapies. Variables were incorporated in a step-wise manner. The ability of each algorithm to identify patients with hemodynamically determined PAH was assessed.

External validation

Data from the University of Virginia Health System was used to externally validate the algorithms. Patients referred to the PH clinic from August 2015 to August 2016 were included, and a similar approach was used as the development algorithms. A cardiologist and a pulmonary physician evaluated all patients jointly in this clinic. Patients were
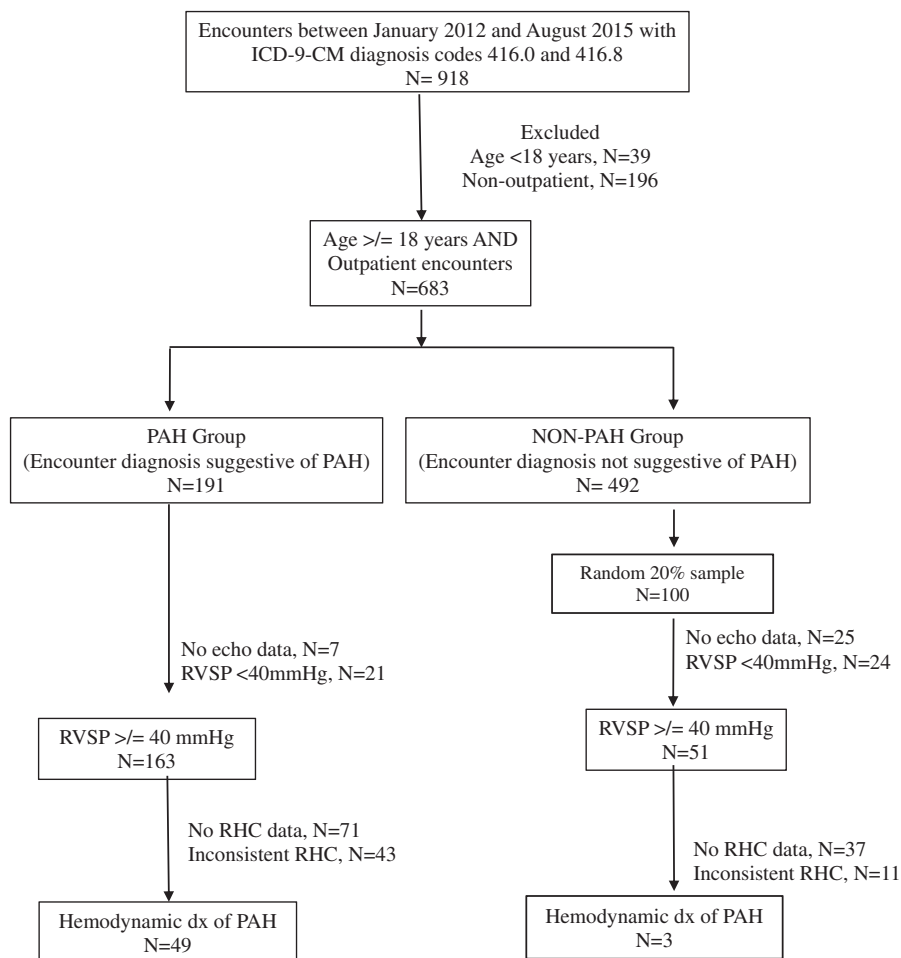
Fig. 1. Development cohort selection and dichotomization based on EMR encounter diagnosis. ICD-9-CM, International Classification of Diseases-9-Clinical Modification; PAH, pulmonary arterial hypertension; RHC, right heart catheterization; RVSP, right ventricular systolic pressure.

Table 1. EMR encounter diagnosis terminology for PAH and non-PAH groups.

PAH group (suggestive of PAH)	Non-PAH group (not suggestive of PAH)
Development cohort	
BMPR2 PAH (pulmonary arterial hypertension)	Hypertensive pulmonary venous disease
Idiopathic PAH (pulmonary arterial hypertension)	Mild pulmonary hypertension
PAH (pulmonary arterial hypertension)	Moderate to severe pulmonary hypertension
Primary pulmonary HTN	Other chronic pulmonary heart disease
Primary pulmonary hypertension	PHT (pulmonary hypertension)
Primary pulmonary hypertensive arterial disease	Pulmonary HTN
Pulmonary arterial hypertension	Pulmonary hypertension
Pulmonary artery hypertension	Pulmonary hypertension, mild
Pulmonary arterial hypertension with portal hypertension	Pulmonary hypertension, moderate to severe
PAH (pulmonary arterial hypertension) with portal hypertension	Pulmonary hypertension, secondary
	Pulmonary hypertensive venous disease
	Pulmonary hypertension with unclear multi-factorial mechanisms
	Secondary pulmonary hypertension
Validation Cohort	
Primary pulmonary hypertension	Other chronic pulmonary hypertension

identified by ICD-9-CM codes and the cohort divided into two groups, PAH and non-PAH, based on an EMR encounter diagnosis suggestive of PAH. The same variables were collected and reviewed with data entry to the algorithms.

Statistical analysis

Descriptive statistics and chi-square tests were used to compare patient demographics and co-morbidities for the PAH and non-PAH groups. Patient co-morbidities were determined using the Elixhauser Comorbidity Index[15] and patient diagnoses at the time of the initial visit. Algorithms were created, with variables included in a stepwise process. For each algorithm, a 2×2 contingency table was created utilizing hemodynamically determined PAH as the gold standard. Patients without an echocardiogram or RHC were allocated to the negative test column for the group. Performance characteristics for each algorithm were calculated, including sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV). Since we reviewed a random sample of only 100 patients from the non-PAH group, we weighted these values by 4.92 in order to return the non-PAH group to the original sample size of 492. Additionally, a logistic regression analysis with the outcome of hemodynamically determined PAH was performed to calculate the odds ratio and c-statistic for each algorithm along with the 95% confidence interval (CI). In each logistic regression model, the predictor variable was defined in an identical manner as the test positive and test negative group in the 2×2 tables for each algorithm. For example, in our algorithm with ICD codes, EMR encounter diagnosis, echocardiogram, and RHC, the indicator would be 1 if each of these was present and 0 if any were absent. All statistical analyses were performed using SAS version 9.4 (SAS Inc., Cary, NC, USA). The level of significance was set at $P \leq 0.05$.

Results

From January 2012 to August 2015, 683 patients with an ICD-9-CM code for PH were identified and based on the EMR encounter diagnosis classified as PAH (n = 191) and non-PAH (n = 492) groups (Fig. 1). Baseline characteristics revealed a similar mean age for both groups: 63.9 ± 15.8 and 64.6 ± 15.7 years, respectively (Table 2). Women were more common in the PAH (71.2%) group compared to the non-PAH (62.8%) group. Diabetes mellitus and chronic diseases of the lung and liver were more frequently observed in the PAH group. Echocardiography results were present in 96% of patients in the PAH group and 74% of patients in the non-PAH group. RHC measurements were performed in 53% of patients in the PAH group and 27% of patients in the non-PAH group. PAH-specific therapy was more frequently prescribed to patients with PAH group (35 %) compared to non-PAH group (12%). A hemodynamic diagnosis

Table 2. Baseline characteristics of the development cohort: patients in PAH and non-PAH groups seen in an outpatient clinic from January 2012 to August 2015.

	PAH (n = 191)	Non-PAH (n = 492)	P value
Age (mean (SD)) (years)	63.88 (15.8)	64.56 (15.7)	0.615
<30	5 (2.6)	15 (3.1)	
31–40	12 (6.3)	28 (5.7)	
41–50	21 (10.9)	52 (10.6)	
51–60	42 (21.9)	101 (20.5)	
61–70	42 (21.9)	103 (20.9)	
71–80	38 (19.9)	116 (23.6)	
81–90	28 (14.7)	68 (13.8)	
90+	3 (1.6)	9 (1.8)	
Sex			0.039
Female	136 (71.2)	309(62.8)	
Male	55 (28.8)	183 (37.2)	
Race			0.565
Not Hispanic or Latino	122 (63.9)	335 (68.1)	
Unknown	37 (19.4)	82 (16.7)	
Hispanic or Latino	32 (16.6)	75 (15.2)	
Co-morbidities			
Hypertension	112 (58.6)	289 (58.7)	0.981
Congestive heart failure	74 (38.7)	160 (32.5)	0.124
Sleep disordered breathing	49 (25.7)	114 (23.2)	0.494
Diabetes mellitus	58 (30.4)	100 (20.3)	0.005
Chronic pulmonary disease	49 (25.7)	90 (18.3)	0.032
Atrial fibrillation	42 (21.9)	89 (18.1)	0.245
Obesity	35 (18.3)	74 (15.1)	0.293
Coronary artery disease	29 (15.2)	72 (14.6)	0.856
Valvular hearth disease	15 (7.9)	51 (10.4)	0.319
Connective tissue disorder	23 (12.0)	46 (9.4)	0.295
Liver disease	16 (8.4)	14 (2.9)	0.002
Atrial flutter	6 (3.1)	7 (1.4)	0.140
Congenital heart disease	2 (1.1)	2 (0.4)	0.312
HIV	3 (1.6)	2 (0.4)	0.136
Interstitial lung disease	0 (0)	2 (0.4)	1.000

of PAH was confirmed in 26% of patients in the PAH group compared with 3% in the non-PAH group.

Development algorithms

Performance characteristics were calculated for eight algorithms in order to identify patients with hemodynamically diagnosed PAH as determined by RHC (Table 3). For claims-based algorithms, sole use of ICD-9-CM codes 416.0 and 416.8 achieved the poorest PPV. Pairing ICD-9-CM codes with a prescription for one PAH-specific medication achieved moderate sensitivity (67.4%), high specificity (86.9%) and high NPV (96.3%), but poor PPV (34.7%).

Table 3. Performance characteristics for claims algorithms in the hemodynamic diagnosis of PAH: Development cohort.

	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	Odds ratio* (95% CI)	C-statistic* (95% CI)
Claims-based algorithms						
ICD-9-CM codes 416.0 and 416.8	–	–	9.34	–		
ICD codes + at least one PAHRx	67.44	86.91	34.67	96.29	13.61 (7.69–24.09)	0.84 (0.79–0.90)
ICD codes + two or more classes PAHRx	28.23	98.56	66.86	93.03	26.87 (11.43–63.14)	0.66 (0.60–0.73)
EMR-based algorithms						
ICD codes + EMR encounter dx	76.85	77.07	25.65	97.00	11.16 (6.08–20.49)	0.67 (0.63–0.72)
ICD codes + EMR encounter dx + echo	76.85	78.20	26.63	97.04	11.91 (6.48–21.89)	0.69 (0.64–0.73)
ICD codes + EMR encounter dx + echo + RHC	76.85	91.44	48.04	97.46	35.38 (18.60–67.32)	0.86 (0.82–0.90)
ICD codes + EMR encounter dx + echo + RHC + PAHRx	67.44	96.93	69.35	96.66	65.52 (32.76–131.08)	0.87 (0.82–0.93)
ICD codes + EMR encounter dx + PAHRx	67.44	96.45	66.15	96.64	56.31 (28.72–110.40)	0.87 (0.81–0.92)

*Odds ratio and C-statistic came from a logistic regression model with the predictor based on the algorithm.

dx, diagnosis; EMR, electronic medical records; RHC, right heart catheterization; PAHRx, PAH-specific therapies; PPV, positive predictive value; NPV, negative predictive value.

Combining ICD-9-CM codes with prescriptions for more than one class of PAH-specific medication improved PPV (66.9%) and specificity (98.6%).

Subsequently, we calculated the performance of EMR-based algorithms that included the ICD-9-CM code, EMR encounter diagnosis, performance of echocardiography, performance of RHC, and prescription of PAH-specific therapy in a step-wise manner. The addition to ICD-9-CM codes of an EMR encounter diagnosis of PAH (Table 3) resulted in a PPV of 25.7%. The addition of echocardiography performance to the algorithm produced minimal improvement in the algorithm performance characteristics. However, the addition of RHC performance increased the PPV (48.0%). The algorithm with the best performance characteristics was observed with a combination of ICD-9-CM codes, EMR encounter diagnosis of PAH, echocardiography, RHC, and a prescription for PAH-specific medication (PPV 69.4%, sensitivity 67.4%). Lastly, the algorithm that contained ICD-9-CM codes, an EMR encounter diagnosis of PAH, and a prescription for PAH-specific medication yielded a modest sensitivity (67.4%) and modest PPV (66.2%). Finally, we calculated odds ratio and the c-statistic using multiple logistic regression model. As shown in Table 3, the performance characteristics of the model to predict PAH was best for combined ICD codes and a prescription of at least one PAH therapy (c-statistic = 0.84, 95% CI = 0.79–0.90). Interestingly, additional variables such as EMR encounter diagnosis, presence of echo and or RHC did not improve the c-statistic.

External validation

External validation was conducted at the University of Virginia Health System that included 177 patients with an ICD-9-CM code for PH (Fig. 1e, available in the online Supplementary Material). Patients were classified into PAH (n = 28) and non-PAH (n = 149) groups based on the EMR encounter diagnosis for at least one clinic visit. The mean age for both groups was similar for PAH and non-PAH groups: 66.4 ± 24.9 and 67.3 ± 15.1 years, respectively (Table 1e, available in the online Supplementary Material). In addition, no difference in gender distribution was noted in PAH and non-PAH groups. Congenital heart disease was more common in the PAH group while congestive heart failure was more often recorded in the non-PAH group. Of note, in the external validation electronic healthcare record, only two EMR encounter diagnoses were used: primary PH (PAH group) and other chronic pulmonary heart disease (non-PAH group). As with the development algorithm, the performance of ICD-9-CM codes alone achieved the poorest PPV (15.8 %) (Table 4). The algorithm that combined ICD-9-CM codes and more than one class of PAH-specific therapies, yielded a higher PPV (57.14%) and specificity (93.96%), similar to development algorithms. Likewise, the EMR-based algorithm with ICD-9-CM codes, EMR encounter diagnosis, performance of echocardiography and RHC, and a prescription for PAH-specific medication yielded the greatest PPV (62.5%). Performance of the development and external validation claims-based algorithms that included ICD-9-CM codes and more than one PAH-specific therapy was similar with respect to the PPV: 66.86% and 57.14%, respectively ($P = 0.39$). The lower sensitivity and a higher proportion of patients with congenital heart disease and CHF were consistent with the referral pattern to this clinic as stated in the method section.

Discussion

In this retrospective cohort study, we developed and attempted to validate the performance of claims-based and EMR-based algorithms to identify patients with PAH as determined by RHC. To our knowledge, this extensive analysis using administrative datasets to distinguish patients with PAH from those with other pulmonary vascular disease

Table 4. Performance characteristics for claims algorithms in the hemodynamic diagnosis of PAH: Validation cohort.

	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)
Claims-based algorithms				
ICD-9-CM codes 416.0 and 416.8	–	–	15.82	–
ICD codes + at least one PAHRx	64.29	81.88	40.00	92.42
ICD codes + more than one PAHRx	42.86	93.96	57.14	89.74
EMR-based algorithms				
ICD codes + EMR encounter dx	25.00	85.91	25.00	85.91
ICD codes + EMR encounter dx + echo	25.00	91.28	35.00	86.62
ICD codes + EMR encounter dx + echo + RHC	25.00	96.64	58.33	87.27
ICD codes + EMR encounter dx + echo + RHC + PAHRx	17.86	97.99	62.50	86.39
ICD codes + EMR encounter dx + PAHRx	17.86	95.97	45.45	86.14

dx, diagnosis; EMR, electronic medical records; RHC, right heart catheterization; PAHRx, PAH-specific therapies; PPV, positive predictive value; NPV, negative predictive value.

states has not been reported. Our results indicate that sole use of a code-based algorithm does not provide accurate estimates of PAH, based on the poor PPV. Consistent with previous reports, ICD-9-CM codes 416.0 and 416.8 alone achieved a low PPV.[10–12] Moreover, in the present study, the performance of algorithms that included diagnostic tests or PAH-specific treatments or both was superior to use of ICD-9-CM codes alone.

Claims-based algorithms were developed and validated for various common health conditions.[16–19] In addition, claims-based studies have been published to describe healthcare utilization, treatment patterns, and mortality from PAH.[9,10,20–24] In an epidemiological study of PAH using Scottish Morbidity Record data, Peacock et al. used primary PH codes (ICD-9-CM code 416.0 and ICD-10 I27.0) with a series of filters applied in the inclusion process.[20] These filters included claims for conditions associated with World Health Organization (WHO) Group II–V Pulmonary Hypertension (left heart disease, lung disease, chronic thromboembolic, or miscellaneous causes). Davis et al. used ICD-9-CM code 416.0 to study PAH-related mortality patterns.[21] Kirson et al. used the same algorithm to study the prevalence and excess costs associated with PAH in a privately insured population.[24,25] In a recent study, PAH-related hospitalization trends and outcomes were studied with Nationwide Inpatient Sample (NIS) data, using ICD-9-CM code 416.0.[9] Although these reports have been instrumental in understanding PAH at the population level, none of the case definitions or claims algorithms used in these reports were validated. In addition, ICD-9-CM codes 416.0 and 416.8 are often used interchangeably during different clinic visits. The addition of claims for echocardiogram, RHC, and introduction of PAH-specific medication improves the PPV of a claims-based algorithm. This reflects a natural workup of a patient with true PAH and can be built using claims history. One drawback is the inability to get actual values of diagnostic procedures from claims. However, the prescription of a PAH-specific therapy in a patient who undergoes these investigations based on

clinician pretest probability suggests that true PAH cases can be mined from claims data and improve the diagnostic accuracy of the algorithm.

The ICD-9-CM codes 416.0 ("primary") and 416.8 ("secondary") do not differentiate PAH from non-PAH patients and they certainly do not reflect the current WHO classification of PH. In addition, the diagnosis of PAH is complex even at the individual patient level and requires consideration of broad differential diagnosis and a multitude of diagnostic tests. Moreover, after a thorough investigation, the diagnosis of PH may change to a different clinical classification group. This complexity highlights the difficulty in isolating PAH patients at the population level. After reviewing three different databases, Link et al. concluded that the fluctuations in PAH-related mortality and hospitalization rates are related merely to coding changes.[12] In a most recent study, Geva et al. concluded that mining EMRs using computable phenotype algorithms identified a large number of pediatric PH patients, who were otherwise not included in the PH registry.[26] The current study reinforces the importance of validation studies when using ICD codes. Once an algorithm is thoroughly tested and validated in various settings, it becomes a tool in defining disease epidemiology and healthcare utilization accurately at the population level.

Our study has several limitations. First, this study is a retrospective analysis of claims data from two referral centers which are familiar with diagnosis, testing, and management of PH patients. These results need to be replicated in other settings like large databases before generalization. Second, patients without echocardiogram and RHC findings were included in the test-negative group of the algorithms and this arrangement might have compromised the performance characteristics of various algorithms, which may perform better in settings in which all diagnostic information is available. Third, we included phosphodiesterase-5 inhibitors in PAH-specific therapies, which likely result in low PPV as this class of drugs can be prescribed for other conditions (i.e. erectile dysfunction). Fourth, the EMR phrases of our

external validation cohort are different from the original cohort and highlight the importance of building a universal data dictionary within the EMR system. However, in our study, the content of the phrases that was used in the development cohort was similar. Fifth, the lower sensitivity of the algorithms in the validation cohort reflects a referral bias. The PH clinic at the validation cohort site was staffed by cardiology and pulmonary physicians that likely resulted in a higher prevalence of pulmonary venous hypertension. While the algorithm that we developed is superior to existing claims-based algorithms, future investigations should validate this algorithm in different settings such as large, urban referral centers and community-based clinics with low disease prevalence. Lastly, adding additional elements identified more patients with PAH without improving the c-statistic. Thus, limiting the algorithm to two elements (ICD-9-CM codes + PAH-specific therapy) would prevent unnecessary classification of patients without the disease in population-based studies.

In conclusion, the use of ICD-9-CM codes alone resulted in poor positive predictive value to identify patients with PAH. However, ICD-9-CM codes and a prescription for PAH-specific medications improved the performance characteristics of the algorithms and could be used for population-based studies.

ORCID iD

Ravikanth Papani ● http://orcid.org/0000-0002-5384-9150

References

1. Simonneau G, Gatzoulis MA, Adatia I, et al. Updated clinical classification of pulmonary hypertension. *J Am Coll Cardiol* 2013; 62: D34–41.
2. Galiè N, Manes A, Negro L, et al. A meta-analysis of randomized controlled trials in pulmonary arterial hypertension. *Eur Heart J* 2009; 30: 394–403.
3. Liu HL, Chen XY, Li JR, et al. Efficacy and safety of pulmonary arterial hypertension-specific therapy in pulmonary arterial hypertension: a meta-analysis of randomized controlled trials. *Chest* 2016; 150: 353–366.
4. Badesch DB, Raskob GE, Elliott CG, et al. Pulmonary arterial hypertension: baseline characteristics from the REVEAL Registry. *Chest* 2010; 137: 376–387.
5. Hurdman J, Condliffe R, Elliot CA, et al. ASPIRE registry: assessing the Spectrum of Pulmonary hypertension Identified at a REferral centre. *Eur Respir J* 2012; 39: 945–955.
6. Sithamparanathan S, Nair A, Thirugnanasothy L, et al. Survival in portopulmonary hypertension: Outcomes of the United Kingdom National Pulmonary Arterial Hypertension Registry. *J Heart Lung Transplant* 2017; 36: 770–779.
7. Deaño RC, Glassner-Kolmin C, Rubenfire M, et al. Referral of patients with pulmonary hypertension diagnoses to tertiary pulmonary hypertension centers: the multicenter RePHerral study. *JAMA Intern Med* 2013; 173: 887–893.
8. McLaughlin VV, Langer A, Tan M, et al. Contemporary trends in the diagnosis and management of pulmonary arterial hypertension: an initiative to close the care gap. *Chest* 2013; 143: 324–332.
9. Anand V, Roy SS, Archer SL, et al. Trends and outcomes of pulmonary arterial hypertension-related hospitalizations in the United States: analysis of the nationwide inpatient sample database from 2001 through 2012. *JAMA Cardiol* 2016; 1: 1021–1029.
10. Hyduk A, Croft JB, Ayala C, et al. Pulmonary hypertension surveillance–United States, 1980–2002. *MMWR Surveill Summ* 2005; 54: 1–28.
11. George MG, Schieb LJ, Ayala C, et al. Pulmonary hypertension surveillance: United States, 2001 to 2010. *Chest* 2014; 146: 476–495.
12. Link J, Glazer C, Torres F, et al. International Classification of Diseases coding changes lead to profound declines in reported idiopathic pulmonary arterial hypertension mortality and hospitalizations: implications for database studies. *Chest* 2011; 139: 497–504.
13. Duarte AG, Lin YL and Sharma G. Incidence of right heart catheterization in patients initiated on pulmonary arterial hypertension therapies: A population-based study. *J Heart Lung Transplant* 2017; 36: 220–226.
14. Fritz JS and Smith KA. The pulmonary hypertension consult: clinical and coding considerations. *Chest* 2016; 150: 705–713.
15. Elixhauser A, Steiner C, Harris DR, et al. Comorbidity measures for use with administrative data. *Med Care* 1998; 36: 8–27.
16. Andrade SE, Harrold LR, Tjia J, et al. A systematic review of validated methods for identifying cerebrovascular accident or transient ischemic attack using administrative data. *Pharmacoepidemiol Drug Saf* 2012; 21(Suppl 1): 100–128.
17. Abraha I, Serraino D, Giovannini G, et al. Validity of ICD-9-CM codes for breast, lung and colorectal cancers in three Italian administrative healthcare databases: a diagnostic accuracy study protocol. *BMJ Open* 2016; 6: e010547.
18. Lacasse Y, Daigle JM, Martin S, et al. Validity of chronic obstructive pulmonary disease diagnoses in a large administrative database. *Can Respir J* 2012; 19: e5–9.
19. Saczynski JS, Andrade SE, Harrold LR, et al. A systematic review of validated methods for identifying heart failure using administrative data. *Pharmacoepidemiol Drug Saf* 2012; 21(Suppl 1): 129–140.
20. Peacock AJ, Murphy NF, McMurray JJ, et al. An epidemiological study of pulmonary arterial hypertension. *Eur Respir J* 2007; 30: 104–109.

21. Davis KK, Lilienfeld DE and Doyle RL. Increased mortality in African Americans with idiopathic pulmonary arterial hypertension. *J Natl Med Assoc* 2008; 100: 69–72.

22. Copher R, Cerulli A, Watkins A, et al. Treatment patterns and healthcare system burden of managed care patients with suspected pulmonary arterial hypertension in the United States. *J Med Econ* 2012; 15: 947–955.

23. Burke JP, Hunsche E, Régulier E, et al. Characterizing pulmonary hypertension-related hospitalization costs among Medicare Advantage or commercially insured patients with pulmonary arterial hypertension: a retrospective database study. *Am J Manag Care* 2015; 21: s47–58.

24. Kirson NY, Birnbaum HG, Ivanova JI, et al. Prevalence of pulmonary arterial hypertension and chronic thromboembolic pulmonary hypertension in the United States. *Curr Med Res Opin* 2011; 27: 1763–1768.

25. Kirson NY, Birnbaum HG, Ivanova JI, et al. Excess costs associated with patients with chronic thromboembolic pulmonary hypertension in a US privately insured population. *Appl Health Econ Health Policy* 2011; 9: 377–387.

26. Geva A, Gronsbell JL, Cai T, et al. A computable phenotype improves cohort ascertainment in a pediatric pulmonary hypertension registry. *J Pediatr* 2017; 188: 224–231.