

Development and Validation of Algorithms to Identify Pulmonary Arterial Hypertension in Administrative Data



Kari R. Gillmeyer, MD; Eduardo R. Nunez, MD; Seppo T. Rinne, MD, PhD; Shirley X. Qian, MS; Elizabeth S. Klings, MD; and Renda Soylemez Wiener, MD, MPH

BACKGROUND: Pulmonary arterial hypertension (PAH) is a rare disease, and much of our understanding stems from single-center studies, which are limited by sample size and generalizability. Administrative data offer an appealing opportunity to inform clinical, research, and quality improvement efforts for PAH. Yet, currently no standardized, validated method exists to distinguish PAH from other subgroups of pulmonary hypertension (PH) within this data source.

RESEARCH QUESTION: Can a collection of algorithms be developed and validated to detect PAH in administrative data in two diverse settings: all Veterans Health Administration (VA) hospitals and Boston Medical Center (BMC), a PAH referral center.

STUDY DESIGN AND METHODS: In each setting, we identified all adult patients with incident PH from 2006 through 2017 using International Classification of Diseases PH diagnosis codes. From this baseline cohort of all PH subgroups, we sequentially applied the following criteria: diagnosis codes for PAH-associated conditions, procedure codes for right heart catheterizations (RHCs), and pharmacy claims for PAH-specific therapy. We then validated each algorithm using a gold standard review of primary clinical data and calculated sensitivity, specificity, positive predictive values (PPVs), and negative predictive values.

RESULTS: From our baseline cohort, we identified 12,012 PH patients in all VA hospitals and 503 patients in BMC. Sole use of PH diagnosis codes performed poorly in identifying PAH (PPV, 16.0% in VA hospitals and 36.0% in BMC). The addition of PAH-associated conditions to the algorithm modestly improved PPV. The best performing algorithm required ICD diagnosis codes, RHC codes, and PAH-specific therapy (VA hospitals: specificity, 97.1%; PPV, 70.0%; BMC: specificity, 95.0%; PPV, 86.0%).

INTERPRETATION: This set of validated algorithms to identify PAH in administrative data can be used by the PAH scientific and clinical community to enhance the reliability and value of research findings, to inform quality improvement initiatives, and ultimately to improve health for PAH patients.

CHEST 2021; 159(5):1986-1994

KEY WORDS: administrative data; medical informatics; pulmonary arterial hypertension; pulmonary hypertension

FOR EDITORIAL COMMENT, SEE PAGE 1713

ABBREVIATIONS: BMC = Boston Medical Center; ICD = International Classification of Diseases; ICD-9 = International Classification of Diseases, Ninth Revision; ICD-10 = International Classification of Diseases, Tenth Revision; PAH = pulmonary arterial hypertension; PH = pulmonary hypertension; PPV = positive predictive value;

RHC = right heart catheterization; VA = Veterans Health Administration; WHO = World Health Organization

AFFILIATIONS: From the Center for Healthcare Organization & Implementation Research (K. R. Gillmeyer, E. R. Nunez, S. T. Rinne, S. X. Qian), Edith Nourse Rogers Veterans Hospital, Bedford, MA;

Pulmonary arterial hypertension (PAH), a rare subgroup of the heterogeneous diagnosis of pulmonary hypertension (PH), is a devastating disease of the pulmonary vasculature that exerts a heavy burden of symptoms on patients and often leads to accelerated mortality.¹⁻³ Because of its rarity, much of what we know of PAH comes from small, single-center cohort studies and registries of PAH patients from expert referral centers.⁴ These studies have been valuable in building our knowledge of PAH, including characterizing patient demographics, determining prognostic factors, and establishing the natural history of the disease. However, these cohorts and registries may not always capture PAH patients managed in the community outside of referral centers.⁵ Indeed, PAH practice patterns in the community often diverge from expert care, with large gaps in the recognition and management of the disease.⁶⁻⁹ Additionally, PAH registries often lack representation of racial and ethnic minority groups, a critical problem recognized by the PAH community.¹⁰

A clear need exists to expand our understanding of PAH beyond registry data to the population level.

Administrative data generated by billing activities and electronic health records provide an opportunity to conduct large-scale epidemiologic studies to capture community PAH care better and to include historically underrepresented racial and ethnic groups. Administrative data also may be used to develop and implement quality improvement initiatives, to inform policy decisions, and to support learning health systems.¹¹ However, a present challenge with using this type of data lies in how PH is labeled in the International Classification of Diseases,

Ninth Revision (ICD-9) and Tenth Revision (ICD-10), coding systems. ICD PH diagnosis codes identify patients as simply having primary or secondary PH and do not mirror the current World Health Organization (WHO) clinical classification,¹² which categorizes PH into five groups according to underlying causes, hemodynamics, and medical management, with PAH considered group 1 PH. Examining PAH independently from other groups of PH is critical because the epidemiologic features, prognosis, treatment options, and health-care resource costs for PAH patients are distinct from those of patients with other groups of PH.¹³ Unfortunately, as we showed in a recent systematic review,¹⁴ oversimplified algorithms solely using ICD PH diagnosis codes perform poorly in differentiating PAH from other, more common groups of PH, with positive predictive values (PPVs) as low as 3%.

To overcome this limitation of the ICD coding system, PH researchers have created nuanced algorithms to distinguish PAH from other groups of PH. These algorithms vary widely with no standardization between studies, and most algorithms have not been validated.¹⁴ In response to this unmet need, a panel of PH clinical and research experts recently convened to create best practices for developing algorithms to identify PAH in administrative data.¹⁵ These recommendations included anticipated performance characteristics of the algorithms, although none of the algorithms were validated directly through primary data. Therefore, we sought to create and validate systematically a collection of algorithms following these expert recommendations that can be used by the PAH research and clinical community to address diverse research questions and to capitalize on the rich data available through administrative sources.

the VA Boston Healthcare System (R. Soylemez Wiener), Boston, MA; and The Pulmonary Center (K. R. Gillmeyer, E. R. Nunez, S. T. Rinne, E. S. Klings, R. Soylemez Wiener), Boston University School of Medicine, Boston, MA.

FUNDING/SUPPORT: This work was supported by the Veterans Health Administration [HSR&D IIR Grant 15-115; PI: R. S. W.], the National Institutes of Health [NRSA Grant 1F32HL149236-01; PI: K. R. G.], and by resources from the Edith Nourse Rogers Memorial Veterans Hospital and VA Boston Healthcare System. In addition, S. T. R. is funded by a Veterans Health Administration HSR&D Career Development Award. E. S. K. is funded by the National Heart, Lung, and Blood Institute [Grants 1U01HL128566 and 1UG3 HL143192-01A1]. The secure web platform used for managing the data abstraction tool, REDCap, was supported by Boston University [CTSI Grant 1UL1TR001430]. Support for VA data provided by the Department of Veterans Affairs, VA Health Services Research and Development Service, VA Information Resource Center (Project Numbers SDR 02-237 and 98-004).

CORRESPONDENCE TO: Kari R. Gillmeyer, MD; e-mail: kgill@bu.edu
Copyright © 2020 American College of Chest Physicians. Published by Elsevier Inc. All rights reserved.

DOI: <https://doi.org/10.1016/j.chest.2020.12.010>

Methods

Study Settings

We performed retrospective studies of adults with PH diagnosed between January 1, 2006, and December 31, 2017, in two distinct settings: the Veterans Health Administration (VA), the largest integrated health-care system in the United States that provides care to veterans in diverse economic and geographic locations across the spectrum of inpatient and outpatient settings, and Boston Medical Center (BMC), a large safety-net hospital and academic PAH expert referral center enriched with racial and ethnic minority groups. The Edith Nourse Rogers Memorial VA Hospital and Boston University Institutional Review Boards approved this study.

Algorithm Development

Within both settings (the VA and BMC), we identified all adult (age ≥ 18 years) patients with PH between 2006 (the first full year of data after

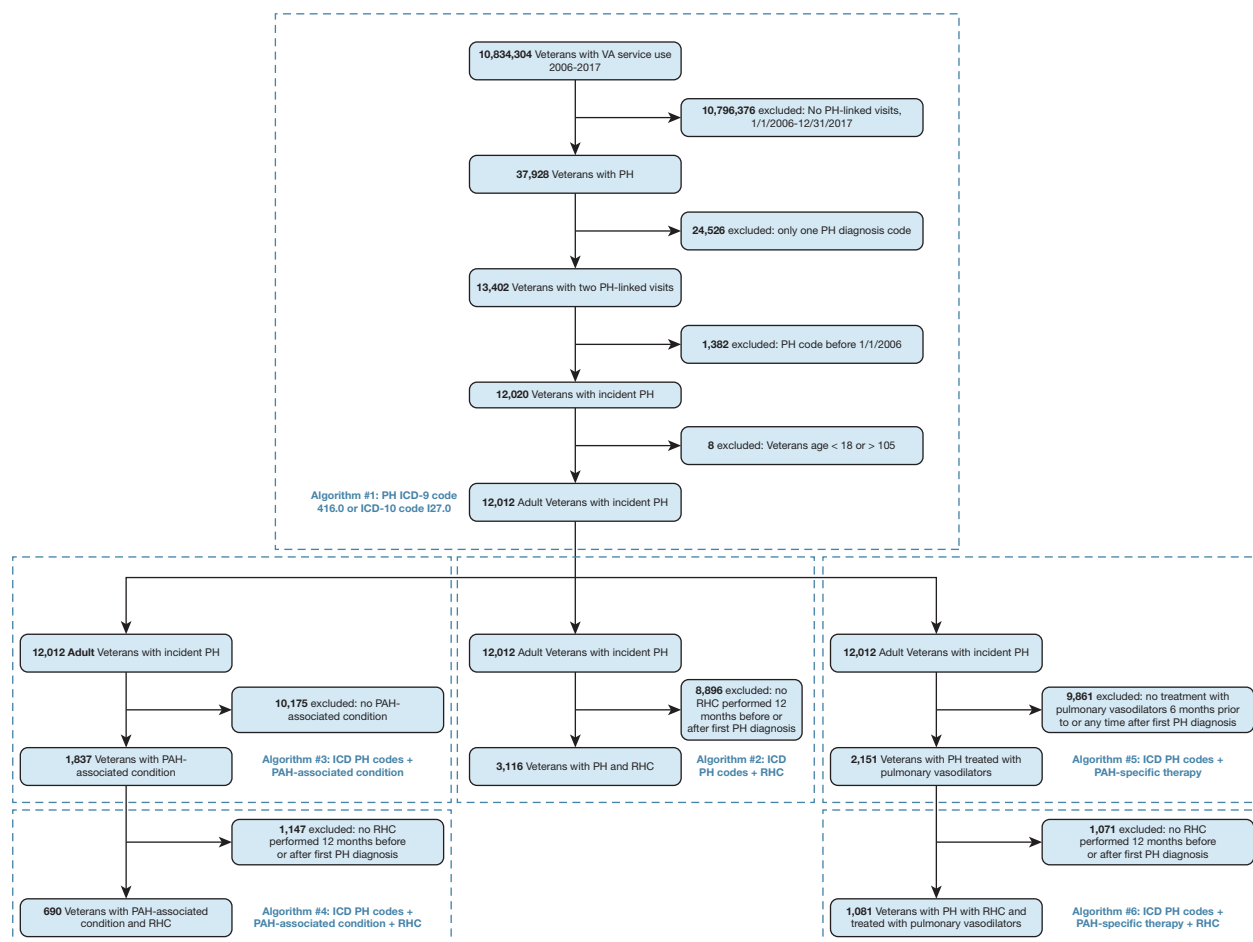


Figure 1 – Algorithm development in the VA setting. ICD = International Classification of Diseases; PAH = pulmonary arterial hypertension; PH = pulmonary hypertension; RHC = right heart catheterization; VA = Veterans Health Administration.

sildenafil, a phosphodiesterase-5-inhibitor, was approved by the Food and Drug Administration for use in PH) and 2017 (most recently available data), defined by at least two visits (either inpatient or outpatient) linked to an ICD-9 (416.0) or ICD-10 (I27.0) diagnosis code for PH. We excluded patients with ICD PH diagnosis codes unlikely to capture PAH, such as ICD-9 code 416.8 (“other chronic pulmonary heart diseases”) and 416.2 (“chronic thromboembolic pulmonary hypertension”). To select incident PH, we excluded those with a PH code before December 31, 2005. From this baseline cohort of all groups of PH, we built a series of algorithms following expert recommendations for PAH algorithm development.¹⁵ We sequentially applied the following criteria: (1) inclusion of ICD diagnosis codes for PAH-associated conditions such as connective tissue disease, congenital heart disease, HIV infection, and portal hypertension; (2) inclusion of ICD or Current Procedural Terminology diagnosis codes for right heart catheterization (RHC), a required diagnostic test to confirm the presence of PAH and to support treatment decisions¹³; and (3) inclusion of pharmacy claims (both inpatient and outpatient) for PAH-specific therapies (prostacyclins, prostacyclin receptor agonists, endothelin receptor antagonists, phosphodiesterase-5-inhibitors, or soluble guanylate cyclase agonists). We considered various combinations of the aforementioned criteria and selected algorithms for chart validation that retained $\geq 5\%$ of the baseline cohort sample to ensure adequate sample size. To ensure that phosphodiesterase-5-inhibitor therapy was intended to treat PAH rather than erectile dysfunction, we

required a dispensing rate of at least 15 pills per month. The derivation of our algorithms within the VA setting is shown in Figure 1. Definitions for each of the above criteria, including ICD and Current Procedural Terminology codes and timelines, are shown in Table 1.

Algorithm Validation

We validated the developed algorithms in each setting using a gold standard review of primary clinical and diagnostic data. For each algorithm created, a pulmonologist (K. R. G.) performed structured chart reviews on a random sample of 50 patients (6 algorithms \times 50 charts = 300 charts each in VA and BMC), extracting the following variables from the chart using a data abstraction tool: (1) demographics, including age, sex, race, and ethnicity; (2) echocardiography results; (3) RHC results, including mean and diastolic pulmonary artery pressure, pulmonary artery wedge pressure, cardiac output, transpulmonary and diastolic pressure gradients, pulmonary vascular resistance, and vasodilator responsiveness, if performed; (4) ancillary testing results, including CT scans of the chest, ventilation-perfusions scans, and pulmonary function testing; and (5) clinic progress notes, consult notes, and discharge summaries. Based on all available information, the presence of PH and WHO PH group was determined. A second pulmonologist (E. R. N.) who was blinded to the first reviewer’s assessment examined a randomly selected 20% of charts reviewed by

TABLE 1] Definitions for Algorithm Criteria

Variable	Definition	ICD-9 Codes	ICD-10 Codes	CPT Codes
PAH-associated condition	ICD code any time before or 12 mo after index date ^a
Connective tissue diseases	ICD code any time before or 12 mo after index date ^a	517.1, 517.2, 695.4, 701.0, 710.0, 710.1, 710.2, 710.3, 710.4, 710.8, 710.9, 714.x	L90.0, L94.0, L93.x, M05.x, M06.x, M08.x, M12.0xx, M32.x, M33.x, M34.x, M35.0x, M35.1, M35.8, M35.9	...
HIV infection	ICD code any time before or 12 mo after index date ^a	042.x, 079.53, V08.x	B20.x, B21.x, B22.x, B23.x, B24.x, B97.35, Z21.x,	...
Portal hypertension	ICD code any time before or 12 mo after index date ^a	572.3	K76.6	...
Congenital heart disease	ICD code any time before or 12 mo after index date ^a	745.x, 747.0, 747.41, 747.42	Q20.0, Q20.1, Q20.2, Q20.3, Q20.4, Q20.5, Q25.0, Q21.x, Q26.2, Q26.3	...
Schistosomiasis	ICD code any time before or 12 mo after index date ^a	120.x	B65.x	...
PCH	ICD code any time before or 12 mo after index date ^a	448.0	I78.0	...
PPHN	ICD code any time before or 12 mo after index date ^a	747.83	P29.30	...
Right heart catheterization	Diagnostic code within 12 mo before or after index date ^a	37.21, 37.23, 89.63, 89.64	4A023N6, 4A023N8, 02HP32Z, 02HQ32Z, 02HR32Z	93451, 93453, 93456, 93457, 93460, 93461, 93463, 93501, 93503, 93526, 93527, 93528, 93529, 93530, 93531, 93532, 93533
PAH-specific therapy	Prescription for pulmonary vasodilator ^b 6 mo before or any time after index date ^a

CPT = Current Procedural Terminology; HIV = human immunodeficiency virus; ICD = International Classification of Diseases; ICD-9 = International Classification of Diseases, Ninth Revision; ICD-10 = International Classification of Diseases, Tenth Revision; PAH = pulmonary arterial hypertension; PCH = pulmonary capillary hemangiomatosis; PPHN = persistent pulmonary hypertension of the newborn.

^aFirst PH diagnosis code.

^bIncluding prostacyclin, prostacyclin receptor agonist, endothelin receptor antagonist, phosphodiesterase-5-inhibitor, or soluble guanylate cyclase agonists. For phosphodiesterase-5-inhibitor therapy, dispensing rate \geq 15 pills/mo.

the first physician from the BMC setting (n = 60) to ensure agreement on PAH diagnosis.^{16,17} Agreement between reviewers regarding presence of PAH was excellent, with a Cohen's κ value of 0.92.¹⁸

Statistical Analysis

For each algorithm, we created a 2 × 2 contingency table using reviewer-determined PAH as the gold standard. Our comparison cohort for each algorithm (those with negative test results) consisted of those from the baseline cohort (those with \geq 2 visit-linked PH diagnosis codes) who did not meet the additional criteria (eg, did not have a PAH-associated condition,

RHC, or PAH-specific therapy, as appropriate). We performed chart reviews on a random subset (n = 50) of each of these comparison cohorts. The choice of these comparison cohorts allowed us to determine the ability of these algorithms to distinguish PAH from other, more common groups of PH. We calculated performance characteristics for each algorithm, including sensitivity, specificity, PPV, and negative predictive value. In line with prior studies, we considered values of \geq 70% to be high, values between 50% and 70% to be moderate or modest, and values < 50% to be poor.¹⁹ An example of a 2 × 2 contingency table is shown in Table 2.

Results

Study Cohort

From our baseline cohort of all groups of incident PH, we identified 12,012 PH patients in the VA and 503 PH patients at BMC. Based on direct chart review, veterans were older (69.5 years vs 59.5 years), predominantly male (90.0% vs 28.0%) and less racially and ethnically diverse compared with BMC patients (Table 3).

Veterans showed a higher prevalence of PH resulting from left heart disease (group 2 PH) and chronic lung disease (group 3 PH) and a lower prevalence of PAH compared with BMC patients. Among patients identified as having PAH through chart review, the most common underlying cause of PAH in both settings was idiopathic PAH (37.2% in the VA and 45.5% in BMC), followed by PAH associated with connective tissue diseases (33.6% in the VA and 38.4% in BMC). Veterans showed a higher burden of PAH secondary to portal hypertension, whereas BMC patients showed a higher burden of PAH associated with HIV infection.

Algorithm Performance in the VA Setting

We found that the sole use of ICD-9 and ICD-10 PH diagnosis codes (algorithm 1) performed poorly in the VA setting, with a PPV of 16.0% (Table 4). Addition of ICD diagnosis codes for PAH-associated conditions (algorithm 3) achieved modest sensitivity (68.4%) with high specificity (91.1%), although the PPV remained poor (48.0%). Likewise, the algorithm requiring an RHC procedure code (algorithm 2) resulted in high sensitivity (84.0%) and specificity (80.0%), although poor PPV (30.0%). The pairing of PAH-associated condition diagnosis codes with RHC procedure codes (algorithm 4) improved the specificity (97.0%) and PPV (52.0%), although this resulted in a loss of sensitivity (44.2%) and sample size. Pairing ICD PH diagnosis codes with

pharmacy claims for PAH-specific therapy (algorithm 5) led to improved performance characteristics, with high specificity (91.5%) and modest PPV (58.0%) and without a significant cost to sensitivity (86.4%) or sample size. The most restrictive criteria requiring ICD PH diagnosis codes, RHC procedures codes, and pharmacy claims for PAH-specific therapy (algorithm 6) achieved a high specificity (97.1%), PPV (70.0%), and sensitivity (77.6%).

Algorithm Performance in PAH Referral Setting

The algorithms consistently achieved higher PPV in the BMC PAH referral center setting, although they maintained a similar pattern of sensitivity and specificity (Table 5). Sole use of ICD diagnosis codes again performed poorly, with a PPV of 36%. Addition of diagnosis codes for PAH-associated conditions modestly improved the PPV (60.0%), although this resulted in lower sensitivity (68.7%). Pairing PAH-associated conditions with RHC procedure codes led to a high PPV (74.0%) and specificity (96.2%), although this resulted in further loss of sensitivity (56.3%) and sample size. The algorithm requiring a pharmacy claim for PAH-specific therapy performed well, with a specificity of 82.9% and PPV of 78.0%. The additional requirement of both RHC procedure codes and PAH-specific therapy improved the performance characteristics even further, with a specificity of 95.0% and PPV of 86.0%, while still maintaining sensitivity (78.9%).

Discussion

In this retrospective study performed in two distinct settings representing diverse patient populations, we created and validated a collection of algorithms to identify PAH and to distinguish it from other groups of PH in administrative data. Overall, the algorithms achieved a

TABLE 2] Example of a 2 × 2 Contingency Table for an Algorithm^a in the VA Setting

Variable	PAH ^b	No PH or Groups 2-5 PH ^b	Total
Positive test ^c	(24/50) × 1,837 = 882	(26/50) × 1,837 = 955	1,837
Negative test ^d	(2/50) × 10,175 = 407	(48/50) × 10,175 = 9,768	10,175
Total	1,289	10,723	12,012

ICD = International Classification of Diseases; PAH = pulmonary arterial hypertension; PH = pulmonary hypertension; VA = Veterans Health Administration.

^aICD PH codes plus PAH-associated condition.

^bAs determined by gold standard chart review of primary clinical data.

^cPatients selected by the above algorithm.

^dPatients selected from the baseline cohort (n = 12,012) who did not meet the additional criteria (ie, patients who did not have a PAH-associated condition).

TABLE 3] Characteristics of Baseline Cohort in the VA vs BMC Setting Determined by Chart Review

Variable	VA Setting (n = 50)	BMC Setting (n = 50)
Age, y	69.5 ± 11.5	59.5 ± 14.8
Female sex	5 (10.0)	36 (72.0)
Race
Black	12 (24.0)	22 (44.0)
White	36 (72.0)	19 (38.0)
Other	2 (4.0)	9 (18.0)
Ethnicity
Hispanic or Latino	4 (8.0)	7 (14.0)
PH group ^a
1 (PAH)	8 (16.0)	18 (36.0)
2	27 (54.0)	25 (50.0)
3	29 (58.0)	8 (16.0)
4	2 (4.0)	3 (6.0)
5	1 (2.0)	4 (8.0)

Data are presented as No. (%) or mean ± SD. BMC = Boston Medical Center; PAH = pulmonary arterial hypertension; PH = pulmonary hypertension; VA = Veterans Health Administration.

^aValues do not sum to 100% because patients can have multiple groups of PH.

higher PPV in the BMC setting compared with the VA setting, which may reflect differences in the underlying patient populations and practice patterns. Because BMC is a PAH expert referral center, its patient population is enriched with patients at risk for PAH. In contrast, veterans carry a high burden of diseases associated with other groups of PH, including heart disease and lung disease.²⁰ Additionally, differences in coding, testing, and treatment practices likely exist among practitioners at BMC compared with those at the VA.

Consistent with prior findings,²¹⁻²³ indiscriminate use of ICD PH diagnosis codes performed poorly with PPV, ranging from 16.0% in the VA setting to 36.0% in the BMC setting, reflecting the incompatibility of ICD PH diagnosis codes with the WHO clinical classification

scheme and the low prevalence of PAH in the general population.⁴ This finding reinforces that ICD PH diagnosis codes should not be used solely to identify PAH in administrative data, and results from studies using this case definition for PAH should be interpreted cautiously.

The addition of RHC procedure codes to the algorithm led to only modest improvements in the PPV, particularly in the VA setting. RHC is required to confirm the diagnosis of PH, particularly PAH,¹³ although it also may be performed for other indications such as monitoring left-sided heart pressures in response to therapy in known group 2 PH. Thus, the performance of an RHC may not be specific to PAH and may capture false-positive results, particularly in settings such as the

TABLE 4] Performance Characteristics for Algorithms in the VA Setting

Variable	Sample Size (No.)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)
ICD-9 code 416.0 or ICD-10 code I27.0	12,012	— ^c	— ^c	16.0	— ^c
ICD PH codes plus RHC	3,116	84.0	80.0	30.0	96.0
ICD PH codes plus PAH-associated condition ^a	1,837	68.4	91.1	48.0	96.0
ICD PH codes plus PAH-associated condition ^a plus RHC	690	44.2	97.0	52.0	96.0
ICD PH codes plus PAH-specific therapy ^b	2,151	86.4	91.5	58.0	98.0
ICD PH codes plus RHC plus PAH-specific therapy ^b	1,081	77.6	97.1	70.0	98.0

ICD = International Classification of Diseases; ICD-9 = International Classification of Diseases, Ninth Revision; ICD-10 = International Classification of Diseases, Tenth Revision; NPV = negative predictive value; PAH = pulmonary arterial hypertension; PPV = positive predictive value; RHC = right heart catheterization; VA = Veterans Health Administration.

^aIncluding connective tissue diseases, congenital heart disease, HIV infection, and portal hypertension.

^bIncluding prostacyclin, prostacyclin receptor agonist, endothelin receptor antagonist, phosphodiesterase-5-inhibitor, or soluble guanylate cyclase agonist.

^cSensitivity, specificity, and NPV could not be calculated as this is the reference group for the other algorithms.

TABLE 5] Performance Characteristics for Algorithms in the BMC Setting

Variable	Sample Size (No.)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)
ICD-9 code 416.0 or ICD-10 code I27.0	503	— ^c	— ^c	36.0	— ^c
ICD PH codes plus RHC	188	86.0	80.7	62.0	96.0
ICD PH codes plus PAH-associated condition ^a	186	68.7	78.2	60.0	84.0
ICD PH codes plus PAH-associated condition ^a plus RHC	61	56.3	96.2	74.0	92.0
ICD PH codes plus PAH-specific therapy ^b	235	92.0	82.9	78.0	94.0
ICD PH codes plus PAH-specific therapy ^b plus RHC	130	78.9	95.0	86.0	92.0

BMC = Boston Medical Center; ICD=International Classification of Diseases; ICD-9 = International Classification of Diseases, Ninth Revision; ICD-10 = International Classification of Diseases, Tenth Revision; NPV = negative predictive value; PAH = pulmonary arterial hypertension; PPV = positive predictive value; RHC = right heart catheterization.

^aIncluding connective tissue diseases, congenital heart disease, HIV infection, and portal hypertension.

^bIncluding prostacyclin, prostacyclin receptor agonist, endothelin receptor antagonist, phosphodiesterase-5-inhibitor, or soluble guanylate cyclase agonist.

^cSensitivity, specificity, and NPV could not be calculated as this is the reference group for the other algorithms.

VA in which the prevalence of group 2 PH is high. The requirement of ICD diagnosis codes for PAH-associated conditions such as connective tissue diseases, HIV infection, and portal hypertension improved specificity and PPV, although this resulted in an overall lower sensitivity. This loss of sensitivity reflects the exclusion of idiopathic PAH from this algorithm, which by definition is not associated with any underlying conditions. Additionally, the presence of a PAH-associated condition does not necessarily indicate the presence of PAH because patients often carry multiple underlying comorbidities associated with more than one group of PH. Although a patient may be at risk for PAH based on their PAH-associated condition, a thorough clinical evaluation ultimately may determine that the patient's PH is classified most accurately under another group. Pairing RHC procedure codes with PAH-associated conditions improved the specificity and PPV beyond the algorithms with the individual components, although at a cost of further loss of sensitivity and sample size.

The best performing algorithm in both settings required pharmacy claims for PAH-specific therapy and RHC procedure codes, achieving a specificity of 97.1% and a PPV of 70.0% in the VA and a specificity of 95.0% and a PPV of 86.0% in BMC, while still maintaining sensitivity and sample size in each setting. It should be noted that this algorithm may still capture off-label use of PAH-specific therapy in PH groups other than PAH. Additionally, PAH patients who did not undergo an RHC during the period specified in the algorithm or who underwent an RHC or PAH-specific therapy in an external health-care system not captured by the available administrative data would be excluded with application of this algorithm and would result in false-negative results.

To our knowledge, only one other study has developed and validated claims-based algorithms systematically to detect PAH in administrative data. In the study by Papani et al,²¹ an algorithm pairing a single ICD-9 PH diagnosis code (416.0 or 416.8) with one class of PAH-specific medications achieved a specificity of 86.9% and a PPV of 34.7% in the development cohort and a specificity of 81.9% and a PPV of 40.0% in the validation cohort. Requiring more than one class of PAH-specific medications improved the specificity and PPV in both the development cohort (specificity, 98.6%; PPV, 66.9%) and validation cohort (specificity, 94.0%; PPV, 57.1%), although with a significant loss of sensitivity. No algorithms without PAH-specific therapy in the case definition were developed. The higher PPV seen in our algorithms compared with those developed by Papani et al²¹ likely reflects our use of only specific ICD PH diagnosis codes (416.0 or I27.0) and perhaps differences in PAH prevalence in the underlying patient populations.

Our collection of algorithms offers PAH clinicians, researchers, and policy makers the ability to address diverse questions while providing increased confidence in the findings. Various factors will need to be considered when deciding on which algorithm to use, including the available administrative data (medical vs pharmacy claims), the outcome to be investigated, the sample size required for the anticipated analyses, and any subpopulation to be examined (eg, patients with scleroderma). As previously discussed,¹⁵ studies aiming to assess the prevalence of PAH, treatment patterns, or cost-effectiveness should use an algorithm with high specificity and PPV, such as the algorithm requiring both PAH-specific therapy and RHC procedure codes. If pharmacy claims are unavailable, or if researchers wish to examine an untreated population, the algorithm using

a PAH-associated condition and RHC procedure code may be a reasonable alternative. The population to which the algorithm will be applied also should be considered. A more restrictive algorithm may be required when examining a population with an expected low underlying prevalence of PAH (ie, community-based settings), whereas a more lenient algorithm may be sufficient when applied to populations with higher prevalence of PAH (ie, PAH referral centers).

Several limitations of these algorithms should be noted. First, although administrative data can confirm if certain diagnostic tests such as a RHC were performed, the granularity of the data does not allow for determination of actual test results, such as the hemodynamic profile obtained during an RHC or lung volumes obtained from pulmonary function tests. Thus, even the most restrictive algorithms will never achieve absolute precision, and the possibility of misclassification will remain. Similarly, because variables such as WHO functional class and hemodynamic severity of RHC are unavailable through these data sources, PAH disease severity cannot be ascertained. Second, application of these algorithms requires an administrative data source to have longitudinal data to establish an appropriate timeline for each algorithm component. Thus, these algorithms may not be applicable to all administrative databases. Likewise, veterans have a unique demographic profile, and thus the performance characteristics seen in the VA cohort may not apply to all community settings, particularly in the underrepresentation of women. Third, although we attempted to exclude group 4 PH patients (chronic thromboembolic pulmonary hypertension) from the algorithms by excluding patients with the associated ICD-9 (416.2) and ICD-10 (I27.82) codes, these patients often undergo RHC and receive treatment with pulmonary vasodilators. Thus, the possibility that some

group 4 PH patients are included in even the most restrictive of algorithms remains. Fourth, although we required an ICD PH diagnosis code for each algorithm in an attempt to maximize specificity and PPV, these algorithms necessarily exclude those without a diagnosis code, which likely reduces the sensitivity of the algorithms. Fifth, the prevalence of PAH in the source population from which the administrative data are derived will impact the PPV of the algorithm, as demonstrated in this study with higher PPV in a PAH referral center (ie, administrative data enriched for PAH will demonstrate better performance characteristics than those enriched for other PH groups). Likewise, when applying these algorithms to a national database such as the VA, the performance characteristics may vary between subsets of the database (ie, some VA medical centers that see high volumes of patients with PAH may have performance characteristics more akin to those seen in the BMC cohort). Finally, although our sample of 50 charts per algorithm was selected randomly, the possibility of sampling error remains, which may affect the precision around the effect estimates.

Interpretation

Accurately distinguishing PAH from other groups of PH in administrative data creates wide opportunities to examine PAH-specific real-world practice patterns, health-care use and costs, quality of care metrics, and patient-centered outcomes and to include underrepresented racial and ethnic groups in research and policy efforts. We developed and validated a collection of algorithms to improve that accuracy and thereby enhance the integrity and value of research findings and quality improvement efforts, to inform policy decisions better, and ultimately to improve the quality of care and health for patients with PAH.

Acknowledgments

Author contributions: K. R. G. had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. K. R. G., S. T. R., E. S. K., and R. S. W. conceived and designed the study. K. R. G., E. R. N., and S. X. Q. acquired the data. All authors analyzed and interpreted the data. K. R. G., S. T. R., E. S. K., and R. S. W. drafted the manuscript. All authors performed critical revision of the manuscript for important intellectual content.

Financial/nonfinancial disclosures: The authors have reported to *CHEST* the following: E. S. K. received research support from Actelion Pharmaceuticals, Pfizer, Bayer, Reata, Micelle, Novartis, and Arena. None declared (K. R. G., E. R. N., S. T. R., S. X. Q., and R. S. W.).

Role of sponsors: The sponsors had no role in the design of the study, the acquisition and analysis of the data, or the drafting of the manuscript.

Other contributions: The views expressed in this article do not necessarily represent the views of the Department of Veterans Affairs or the United States Government.

References

1. Gu S, Hu H, Dong H. Systematic review of health-related quality of life in patients with pulmonary arterial hypertension. *Pharmacoeconomics*. 2016;34:751-770.
2. Delcroix M, Howard L. Pulmonary arterial hypertension: the burden of disease and impact on quality of life. *Eur Respir Rev*. 2015;24:621-629.
3. Benza RL, Miller DP, Barst RJ, Badesch DB, Frost AE, McGoon MD. An evaluation of long-term survival from time of diagnosis in pulmonary arterial hypertension from the REVEAL Registry. *Chest*. 2012;142:448-456.
4. McGoon MD, Benza RL, Escribano-Subias P, et al. Pulmonary arterial hypertension: epidemiology and registries. *J Am Coll Cardiol*. 2013;62:D51-D59.
5. Preston IR, Hinzmann B, Heinz S, et al. An international physician survey of pulmonary arterial hypertension management. *Pulm Circ*. 2016;6:338-346.
6. Maron BA, Choudhary G, Khan UA, et al. Clinical profile and underdiagnosis of pulmonary hypertension in US veteran patients. *Circ Heart Fail*. 2013;6:906-912.
7. Deano 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. Duarte AG, Lin YL, 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.
10. Talwar A, Garcia JGN, Tsai H, et al. Health disparities in patients with pulmonary arterial hypertension: a blueprint for action. An Official American Thoracic Society Statement. *Am J Respir Crit Care Med*. 2017;196:e32-e47.
11. Smith MA, Vaughan-Sarrazin MS, Yu M, et al. The importance of health insurance claims data in creating learning health systems: evaluating care for high-need high-cost patients using the National Patient-Centered Clinical Research Network (PCORNet). *J Am Med Inform Assoc*. 2019;26:1305-1313.
12. Simonneau G, Montani D, Celermajer DS, et al. Haemodynamic definitions and updated clinical classification of pulmonary hypertension. *Eur Respir J*. 2019;53(1):1801913.
13. Galiè N, Humbert M, Vachiery JL, et al. 2015 ESC/ERS Guidelines for the diagnosis and treatment of pulmonary hypertension: the Joint Task Force for the Diagnosis and Treatment of Pulmonary Hypertension of the European Society of Cardiology (ESC) and the European Respiratory Society (ERS): Endorsed by: Association for European Paediatric and Congenital Cardiology (AEPC), International Society for Heart and Lung Transplantation (ISHLT). *Eur Respir J*. 2015;46:903-975.
14. Gillmeyer KR, Lee MM, Link AP, Klings ES, Rinne ST, Wiener RS. Accuracy of algorithms to identify pulmonary arterial hypertension in administrative data: a systematic review. *Chest*. 2019;155:680-688.
15. Mathai SC, Hemnes AR, Manaker S, et al. Identifying patients with pulmonary arterial hypertension using administrative claims algorithms. *Ann Am Thorac Soc*. 2019;16:797-806.
16. Quan H, Khan N, Hemmelgarn BR, et al. Validation of a case definition to define hypertension using administrative data. *Hypertension*. 2009;54:1423-1428.
17. Walsh SLE, Wells AU, Desai SR, et al. Multicentre evaluation of multidisciplinary team meeting agreement on diagnosis in diffuse parenchymal lung disease: a case-cohort study. *Lancet Respir Med*. 2016;4:557-565.
18. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33:159-174.
19. Carnahan RM, Moores KG. Mini-Sentinel's systematic reviews of validated methods for identifying health outcomes using administrative and claims data: methods and lessons learned. *Pharmacoepidemiol Drug Saf*. 2012;21(suppl 1):82-89.
20. Yu W, Ravelo A, Wagner TH, et al. Prevalence and costs of chronic conditions in the VA health care system. *Med Care Res Rev*. 2003;60:146S-167S.
21. Papani R, Sharma G, Agarwal A, et al. Validation of claims-based algorithms for pulmonary arterial hypertension. *Pulm Circ*. 2018;8(2):2045894018759246.
22. Fox BD, Azoulay L, Dell'Aniello S, et al. The use of antidepressants and the risk of idiopathic pulmonary arterial hypertension. *Can J Cardiol*. 2014;30:1633-1639.
23. Link J, Glazer C, Torres F, Chin K. 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.