Identifying Patients with Pulmonary Arterial Hypertension Using Administrative Claims Algorithms

Stephen C. Mathai[1]*, Anna Ryan Hemnes[2]*, Scott Manaker[3], Rebekah H. Anguiano[4], Bonnie B. Dean[5], Vishal Saundankar[5], Peter Classi[6], Andrew C. Nelsen[6], Kathryn Gordon[6], and Corey E. Ventetuolo[7]

[1]Johns Hopkins University School of Medicine, Baltimore, Maryland; [2]Department of Medicine, Vanderbilt University, Nashville, Tennessee; [3]University of Pennsylvania Perelman School of Medicine, Philadelphia, Pennsylvania; [4]University of Illinois at Chicago College of Pharmacy, Chicago, Illinois; [5]Xcenda, LLC, Palm Harbor, Florida; [6]United Therapeutics Corporation, Research Triangle Park, North Carolina; and [7]Alpert Medical School of Brown University; Brown University School of Public Health, Providence, Rhode Island

ORCID IDs: 0000-0001-5956-7026 (A.C.N.); 0000-0002-4223-4775 (C.E.V.).

Abstract

Retrospective administrative claims database studies provide real-world evidence about treatment patterns, healthcare resource use, and costs for patients and are increasingly used to inform policy-making, drug formulary, and regulatory decisions. However, there is no standard methodology to identify patients with pulmonary arterial hypertension (PAH) from administrative claims data. Given the number of approved drugs now available for patients with PAH, the cost of PAH treatments, and the significant healthcare resource use associated with the care of patients with PAH, there is a considerable need to develop an evidence-based and systematic approach to accurately identify these patients in claims databases. A panel of pulmonary hypertension clinical experts and researchers experienced in retrospective claims database studies convened to review relevant literature and recommend best practices for developing algorithms to identify patients with PAH in administrative claims databases specific to a particular research hypothesis.

Keywords: pulmonary hypertension; pulmonary arterial hypertension; administrative claims

(Received in original form October 12, 2018; accepted in final form March 13, 2019)

*Co–first authors.

United Therapeutics Corporation convened the expert panel that participated in this study. B.B.D. and V.S. are employees of Xcenda, which received financial support from United Therapeutics Corporation for this project.

Correspondence and requests for reprints should be addressed to Peter Classi, M.Sc., M.B.A., United Therapeutics Corporation, 55 TW Alexander Drive, Research Triangle Park, NC 27709. E-mail: pclassi@unither.com.

CME will be available for this article at www.atsjournals.org.

Ann Am Thorac Soc Vol 16, No 7, pp 797–806, Jul 2019

DOI: 10.1513/AnnalsATS.201810-672CME
Internet address: www.atsjournals.org

Pulmonary hypertension (PH) refers to an increase in mean pulmonary arterial pressure at rest as assessed by right heart catheterization (RHC) (1). The World Health Organization (WHO) clinically classifies PH into five groups according to clinical and pathophysiologic features, including the underlying cause of disease, clinical and hemodynamic characteristics, and medical management (2, 3). Hemodynamic assessments of PH include measuring pulmonary arterial pressure, pulmonary artery wedge pressure (PAWP), cardiac output, diastolic pressure gradient, and pulmonary vascular resistance (1, 3). PH classifications and hemodynamic assessments are used by PH specialists to diagnose and medically manage the condition, by insurers to verify therapeutic appropriateness, and by the Food and Drug Administration (FDA) for approval of new molecular entities or new indications for existing PH treatments (2, 3). Clinically, PH is often grouped as precapillary and postcapillary, distinguished by PAWP less than or equal to 15 mm Hg (precapillary) and PAWP greater than 15 mm Hg (postcapillary). Precapillary PH includes WHO group 1 (pulmonary arterial hypertension [PAH]), group 3 (PH due to lung diseases and/or hypoxia), group 4 (chronic thromboembolic PH), and some group 5 (PH with unclear and/or multifactorial mechanisms). Postcapillary PH corresponds to WHO group 2 (PH due to left heart diseases) (4).

In addition to precapillary hemodynamic findings, PAH (group 1) is characterized by pulmonary vascular resistance greater than 3 Wood units in the absence of other causes of precapillary PH, such as PH due to lung diseases, chronic thromboembolic pulmonary hypertension (CTEPH), or other diseases. Between 10 and 15 people per 1 million population in the United States are diagnosed with PAH

each year (5). PAH includes different endophenotypes that share a similar clinical picture and virtually identical pathological changes of the lung microcirculation (3). RHC remains essential for the diagnosis of PAH (1, 6). Because other groups of PH, such as PH due to lung diseases and CTEPH, look the same hemodynamically, the hemodynamic assessment profile is used in combination with other clinical data to confirm WHO group 1 PAH. This differentiation is commonly achieved using key diagnostic testing, including echocardiography, ventilation–perfusion ($\dot{V}/\dot{Q}$) lung scan, pulmonary function testing, and other studies, as well as clinical history and physical examination. Distinguishing patients with WHO group 1 PAH from patients with WHO groups 2 to 5 is important for disease management, to achieve therapeutic benefit, and to reduce the potentially harmful effects of using targeted treatment in those who are not in WHO group 1. Currently, WHO group 1 PAH and nonoperative group 4 CTEPH are the only groups with FDA-approved therapies.

In recent years, more than 10 drugs have been approved for the treatment of PAH. Thus, there is a need to understand how these newer treatments are being prescribed and used in a real-world setting. A common source to generate this type of real-world evidence is healthcare administrative claims data. In medical practice, every patient encounter with the healthcare system prompts generation of administrative data, which are used for record keeping and billing purposes (7). Each encounter typically was historically assigned an *International Classification of Diseases, Ninth Revision, Clinical Modification* (ICD-9-CM) code. As of October 2015, the *International Classification of Diseases, 10th Revision, Clinical Modification* (ICD-10-CM) diagnosis codes replaced the ICD-9-CM codes for billing purposes. Researchers generally rely on ICD-9-CM and ICD-10-CM diagnosis codes to identify patients. Although typically collected for billing purposes, claims data can provide insights into real-world health outcomes, treatment patterns, healthcare resource use, and associated costs.

The Unmet Need in PH Claims-based Research

Data generated by claims analyses are receiving more emphasis and are being used for making important healthcare and regulatory decisions (8, 9). The 21st Century Cures Act (Pub. L. 114-255, 130 Stat. 1033) requires the FDA to develop a framework and guidance to evaluate real-world evidence to support approvals of new indications for previously approved drugs and to support or fulfill postapproval safety or efficacy study requirements. Health insurance payers also rely on real-world evidence for making formulary and coverage decisions.

Researchers must rely on selective data available in claims. The ICD-9-CM and ICD-10-CM codes for PH in administrative data do not align directly with the five WHO clinical classification groups. In addition, claims data contain procedure codes to indicate if a patient received RHC or had an echocardiogram or $\dot{V}/\dot{Q}$ scan (i.e., to exclude CTEPH), but there are no results of these procedures to provide evidence of a diagnosis. Taken together, no clear standard methodology distinguishes the clinical classifications of PH or identifies the subset of patients with PAH from administrative claims data. Although the ICD-10-CM update attempts to address this issue, the historical issue remains, and these changes will not be useful for claims-based research for years. The ICD-10-CM codes for PH were most recently updated in October 2017.

Despite a growing number of available therapies, PAH is an orphan disease (10), and hence the sample size for any claims-based study remains a major challenge. Researchers must balance between sample size, sensitivity, and specificity when conducting these studies and use other patient-related information to identify patients with PAH. Thus, the potential misclassification of patients with PAH is a limitation of any claims-based assessment. Given these considerations, we sought to provide guidance based on available literature and insights from PH clinical experts and researchers experienced in retrospective claims database studies. The algorithm recommendations can be applied to address different types of research questions about pulmonary vascular disease when administrative claims data are leveraged.

How Is PAH Currently Being Identified in Claims-based Research?

To support algorithm recommendations, we searched literature from 2008 through February 2018 to identify U.S.-based studies that used claims data to study patients with PAH. A total of 18 claims-based studies were identified (Table 1). Three components were commonly used for identifying patients with PAH: diagnosis codes, PAH-specific medications, and performance of RHC or echocardiography (Table 2). Most claims database studies required at least two of these components for identifying patients with PAH; very few relied on only one component for identifying patients with PAH (11–14). All reviewed studies used ICD-9-CM codes. ICD-10-CM codes were not used in the reviewed studies, because the studies were conducted on claims data collected before October 2015. ICD-10-CM codes would be necessary for any claims-based studies conducted thereafter, and suggested ICD-10-CM codes for the diagnosis of PAH are provided in Table 3 (15). RHC and echocardiography were identified using Current Procedural Terminology (CPT) codes and ICD-9-CM procedure codes.

Algorithms Used for Identifying Patients with PAH

The algorithms used in the published literature for identifying patients with PAH varied considerably across the studies. The choice of algorithm was based on the data available (medical and pharmacy claims), subpopulation focus (e.g., patients with human immunodeficiency virus), and study goals. The most restrictive algorithms required use of PH-related diagnosis codes together with PAH-specific medications and RHC procedure codes (16, 17). The least restrictive algorithms required use of only one component: PH-related diagnosis codes or PAH-specific medications (11–14). It should be noted that the choice of PAH algorithm components affects the sensitivity, specificity, and positive predictive value (PPV) of the algorithm. A given research question should influence the employed algorithm with these performance characteristics in mind.

Two studies compared multiple algorithms using administrative databases or electronic medical record (EMR) data. Burger and colleagues (16) used a two-step approach to ascertain which patients had PAH. In the first step, patients starting prostacyclin therapy were selected if they had *1*) a diagnosis of PH or nonspecific

Table 1. Studies using claims or registry data to identify patients with pulmonary arterial hypertension

Reference	Data Source	Study Period	Diagnosis Codes	Procedure Codes	Medications	Other Considerations
Retrospective case–control design						
19	Multiemployer database	January 2002–December 2007	⩾2 Claims with diagnosis of primary PH (ICD-9-CM 416.0) and no claims associated with categories 2–5 of the Venice classification	⩾1 Claim for RHC or echocardiogram	Not used	None to report
21	Truven Health MarketScan Databases	January 2004–December 2009	⩾2 Claims with diagnosis of primary PH (ICD-9-CM 416.0)	⩾1 Claim for RHC or echocardiogram	Not used	None to report
25	Truven Health MarketScan Databases	January 2003–December 2009	⩾1 Inpatient claim or ⩾2 outpatient claims with diagnosis of PH (ICD-9-CM 416.0 or 416.8)	Not used	Required to have pharmacy claim for sildenafil	Study conducted in patients with CTDs (ICD-9-CM)
Retrospective cohort design						
26	Large managed healthcare plan data (commercial, Medicare Advantage, Medicare part D)	January 2006–December 2008	⩾2 Claims with diagnosis of PH (ICD-9-CM 416.0, 416.8 or 416.9)	Not used	Regardless of PAH diagnosis, ⩾1 claim for ERAs or PAs	None to report
20	Private insurance claims database, random sample of Medicare population (5% national sample)	January 1999–December 2007	⩾2 Claims with diagnosis of primary PH (ICD-9-CM 416.0) and no claims associated with categories 2–5 of the Venice and Dana Point classifications	⩾1 Claim for RHC or echocardiogram	Not used	None to report
27	Truven Health MarketScan Databases	January 2005–September 2008	⩾1 Inpatient claim or ⩾2 outpatient claims with diagnosis of PH (ICD-9-CM 416.0 or 416.8)	Not used	⩾1 Claim for Revatio	Receipt of Viagra excluded
23	Optum Research Database	January 2004–December 2008	⩾1 Medical claim with diagnosis of PH (ICD-9-CM 416.0 or 416.8) and PH-related inpatient stay	Claim for RHC (ICD-9-CM procedure code 37.21 or 37.23; CPT codes 93501, 93526–93529)*	⩾1 Claim for ambrisentan, bosentan, i.v. epoprostenol, iloprost, i.v. or s.c. treprostinil, or sildenafil (except Viagra)	None to report
13	Pharmacy claims data from Medco Health Solutions	January 2008–December 2010	Factors associated with adherence to PDE5 inhibitors in the management of PAH were studied.	Not used	Adcirca or Revatio	None to report

(Continued)

Table 1. (*Continued*)

Reference	Data Source	Study Period	Diagnosis Codes	Procedure Codes	Medications	Other Considerations
28	Administrative claims of a large national managed care organization	January 2004–June 2010	⩾1 Claim with diagnosis of primary PH or other chronic pulmonary heart diseases (ICD-9-CM 416.0 or 416.8) or a diagnosis code for a PAH-associated condition AND ⩾2 claims with primary PH diagnosis, ⩾2 claims with PAH-related diagnosis, and ⩾1 claim with PAH-related medication	Not used	⩾1 Claim for bosentan, ambrisentan, tadalafil, sildenafil, iloprost, treprostinil, epoprostenol*	None to report
29	Optum Research Database	January 2007–October 2011	⩾1 Medical claim with diagnosis of PH (ICD-9-CM 416.0 or 416.8)	Not used	⩾1 Claim for sildenafil, tadalafil, iloprost, bosentan, ambrisentan, epoprostenol treprostinil	Viagra (sildenafil) and Cialis (tadalafil) fills had to have daily doses at levels approved for PAH
14	Kaiser Permanente Southern California Health Plan Data	January 2004–November 2012	⩾1 Claim with diagnosis of PH (ICD-9-CM 416.0 or 416.8)	Not used	Not used	Required to have ICD-9-CM diagnosis code for psoriasis or psoriatic arthritis
18	Optum Research Database	January 2002–December 2011	⩾1 Outpatient visit with diagnosis of PH (ICD-9-CM 416.0 or 416.8)	Outpatient visit (CPT 99201–99205 and 99211–99215) Outpatient consultation (CPT 99241–99245)	⩾1 Claim for ambrisentan, bosentan, sildenafil, tadalafil, epoprostenol, iloprost, treprostinil	ICD-9-CM codes for erectile dysfunction were excluded
30	Humana Research Database	January 2009–June 2014	⩾1 Medical claim with diagnosis of PH (ICD-9-CM 416.0, 416.8, or 416.9)	⩾1 Medical claim with cardiac catheterization CPT codes	⩾1 Claim for ERA, PDE5 inhibitor, prostacyclin (diagnosis or procedure codes required)	None to report
12	Optum Research Database	April 2001–December 2012	⩾1 Claim with diagnosis of PH (ICD-9-CM 416.0 or 416.8)	Not used	Not used	Required to have ICD-9-CM diagnosis code for multiple sclerosis or hepatitis C before interferon treatment
16	Truven Health MarketScan Databases	January 2010–December 2015	⩾1 Claim with diagnosis of PH (ICD-9-CM 416.0, 416.8, or 416.9)	⩾1 Claim for RHC, left heart catheterization, cardiac surgery, echocardiogram, or lung/heart-lung transplant)*	⩾1 Claim for prostacyclin, prostacyclin receptor agonist, ERA, PDE5 inhibitor, sGC*	None to report
22	Truven Health MarketScan Databases	January 2010–December 2014	⩾2 Medical claims with diagnosis of PH (ICD-9-CM 416.0 or 416.8) or ⩾2 medical claims with PAH-related condition diagnosis (HIV, portal hypertension, CTD, congenital heart disease)	Not used	⩾1 Outpatient pharmacy claim for ERA, sGC, or PDE5 inhibitor	Receipt of prostacyclin excluded

(*Continued*)

Table 1. (*Continued*)

Reference	Data Source	Study Period	Diagnosis Codes	Procedure Codes	Medications	Other Considerations
11	Truven Health MarketScan Databases	January 2003–December 2014	⩾2 Nondiagnostic claims with diagnosis of PH (ICD-9-CM 416.0)	Not used	Not used	Required to have nondiagnostic claims for SSc
17	EMR data at outpatient clinics at the University of Texas Medical Branch and the University of Virginia	January 2012–August 2015	Diagnosis of PH (ICD-9-CM 416.0 or 416.8)	Not used	PAH-specific medications used in algorithm but not listed in article	None to report

Definition of abbreviations: CPT = Current Procedural Terminology; CTD = connective tissue disease; EMR = electronic medical records; ERA = endothelin receptor antagonist; HIV = human immunodeficiency virus; ICD-9-CM = *International Classification of Diseases, Ninth Revision, Clinical Modification*; i.v. = intravenous; PA = prostaglandin analog; PAH = pulmonary arterial hypertension; PDE5 = phosphodiesterase type 5; PH = pulmonary hypertension; RHC = right heart catheterization; s.c. = subcutaneous; sGC = soluble guanylate cyclase; SSc = systemic sclerosis.
*Selection criteria were discrete (e.g., either diagnosis codes, procedure codes, or medications were used in the selection process, but each was not necessarily required).

pulmonary heart disease, *2*) a PH-related procedure, or *3*) PAH-specific medication. A total of 13,633 patients were identified in the first step. In the second step, patient counts were ascertained for multiple algorithms created by a combination of any two or three of the preceding components. In the algorithms that required an inpatient claim with PH diagnosis or RHC procedure, only half of the patients identified in step 1 were retained. These requirements help to restrict the study population to the subset of patients who are most likely to have PAH. Using the algorithm requiring an outpatient claim with a diagnosis of PH plus a claim for a PAH-related medication resulted in 90% of patients identified in step 1 being retained. Algorithms that modify the criteria to require two outpatient claims or two PAH-specific medication claims resulted in a nearly identical number of patients as obtained with similar algorithms requiring only one claim. Algorithms that required inpatient claims together with other criteria yielded much smaller sample sizes.

A study by Papani and colleagues (17) validated several algorithms for use in claims data against a hemodynamic diagnosis of PAH determined by RHC through EMR data. Claims-based algorithms that included diagnosis codes plus PAH-specific medications better identified patients with PAH than diagnosis codes alone. Requiring two or more classes of PAH-related therapies increased specificity but decreased sensitivity of the algorithm (i.e., identified fewer patients who truly have PAH but reduced the false-positive results). Algorithms using diagnosis codes together with a claim for at least one PAH-related medication have the highest sensitivity and moderate specificity. Using this algorithm improves the likelihood of identifying more patients with PAH, but at the risk of obtaining a few more false–positive results. However, this algorithm precludes the ability to study the untreated population.

Perspectives on the Development and Use of PAH Algorithms

Many algorithms can be useful for assessing healthcare resource use and cost, treatment patterns (including the diagnostic process and PAH-specific drug use), PH and PAH prevalence, or risk of developing PAH. Some

Table 2. Components of pulmonary arterial hypertension definition used in published literature

Components of PAH Algorithm	Criteria	References
Diagnosis codes	Diagnosis codes were used in all studies identified.	11–14, 16–23, 25–30
	ICD-9-CM codes included 416.0 (primary pulmonary hypertension), 416.8 (other chronic pulmonary heart diseases), and 416.9 (chronic pulmonary heart disease, unspecified).	
	ICD-10-CM codes were not used in the reviewed studies, because the studies were conducted using claims data collected before October 2015.	
	ICD-10-CM codes would be necessary for any claims-based studies conducted thereafter.	
	All studies used outpatient claims in the algorithm; inpatient claims were used less frequently.	22, 23, 25, 27
	Use of diagnosis codes for PAH-related conditions (group 1 conditions such as connective tissue disorders) for patient inclusion was not very common.	11, 12, 14, 16, 25, 28
	Some studies used PH-related conditions in groups 2–5 (such as chronic obstructive pulmonary disease or chronic thromboembolic disease) to exclude patients.	19–21
	Patients with erectile dysfunction were excluded to avoid inclusion of patients prescribed PAH-related medication for conditions other than PH.	18, 23
Procedure codes	All studies using RHC also required a diagnosis code for PH.	19–23, 30
	All studies using echocardiography also required a diagnosis code for PH.	19–22
	Some studies using RHC or echocardiography also required a diagnosis code for a condition associated with WHO classification group 1 PAH condition (e.g., HIV, connective tissue disorders).	11, 12, 14
	Timing of RHC procedure and PH diagnosis was considered. Studies required claims for RHC and PH diagnosis to occur between as few as 60 d and as many as 12 mo of each other.	19–23
PAH-specific medication	PAH-related medication classes included ERAs, PDE5 inhibitors, PAs, prostacyclin receptor agonists, and sGC stimulators.	12, 13, 17, 18, 22, 23, 25–30
	Some drugs excluded from these algorithms were CCBs and sildenafil brand Viagra, a PDE5 inhibitor.	23

Definition of abbreviations: CCB = calcium channel blocker; ERA = endothelin receptor antagonist; HIV = human immunodeficiency virus; ICD-9-CM = *International Classification of Diseases, Ninth Revision, Clinical Modification*; ICD-10-CM = *International Classification of Diseases, 10th Revision, Clinical Modification*; PA = prostaglandin analog; PAH = pulmonary arterial hypertension; PDE5 = phosphodiesterase type 5; PH = pulmonary hypertension; RHC = right heart catheterization; sGC = soluble guanylate cyclase; WHO = World Health Organization.

algorithms are better than others for specific outcomes or assessments (Figure 1).

Three-Component Algorithms

Algorithms requiring all three components (i.e., diagnosis codes, PAH-specific medications, and RHC or echocardiography) are the most stringent and would likely maximize PPV and therefore the probability that subjects truly have PAH (16, 17). However, patients with PAH who are not hemodynamically diagnosed within the study period may be misclassified (i.e., may be falsely classified as not having PAH because they are missing the hemodynamic procedure). Although requiring all three components may reduce the sample size, recent studies suggest that a large proportion of patients have received either RHC or echocardiography. Duarte and colleagues (18) found that 42.7% of patients received RHC and 60.2% received any cardiac catheterization during the 12 months before or 15 months after PH diagnosis, and 91.4% of patients taking PAH-related medication had an echocardiogram in the 12 months before or after their diagnosis. Similarly, Fischer and colleagues (11) found that among incident patients with systemic sclerosis (SSc)-related PAH, 82.5% had an echocardiogram and 28.5% had an RHC procedure within 12 months of first diagnosis of SSc.

This three-component algorithm is likely the best method for measuring prevalence of treated patients with PAH in claims databases, with the limitation that this algorithm will miss patients who are either untreated or do not undergo RHC or echocardiography during the study period. This algorithm may also be most appropriate for use in health economic evaluations such as cost-effectiveness models, given that the algorithm will capture a broader spectrum of healthcare resource use for all patients in the study, including diagnostic tests, pharmaceutical treatment patterns, and PAH-related inpatient and outpatient visits. To increase specificity, patients with diagnosis codes for PH-related conditions listed under WHO groups 2 to 5 can be excluded (19–21). However, this algorithm will fail to capture off-label uses of PAH therapy in non–WHO group 1 conditions and patients with PAH and a concomitant group 2 to 5 comorbidity that is not the underlying cause of their PH. When RHC is included, the timing of and order between claims is important. For example, it must be confirmed that the RHC is preceded or followed by a diagnosis of PH or PAH within a specified time period. The RHC should be conducted within 12 months

Table 3. List of ICD-10-CM codes for identifying patients with pulmonary arterial hypertension

PH Clinical Classifications WHO Group 1: PAH	ICD-10-CM
Idiopathic PAH	I27.0
Heritable PAH	
BMPR2	I27.0
ALK-1, ENG, SMAD9, CAV1, KCNK3	I27.0
Unknown	I27.0
Drug and toxin induced	I27.21
Adverse effect of appetite depressants	I27.21 and T50.5X50
Associated with	
Connective tissue disease	I27.21
Systemic sclerosis	I27.21 and M34.x
Systemic lupus erythematosus	I27.21 and M32.x
Other connective tissue diseases	I27.21 and other required ICD-10-CM codes
HIV infection	I27.21
HIV disease	I27.21 and B20
Portal hypertension	I27.21 and K76.6
Congenital heart diseases	I27.9 or 127.89
Eisenmenger syndrome	I27.83
Atrial septal defect	(I27.9 or 127.89) and Q21.1
Ventricular septal defect	(I27.9 or 127.89) and Q21.0
Schistosomiasis	I27.21 and B65.x
Pulmonary veno-occlusive disease and pulmonary capillary hemangiomatosis	I27.29
Persistent pulmonary hypertension of the newborn	I27.20 or I27.89

Definition of abbreviations: ALK-1 = activin type I receptor kinase-like gene; BMPR2 = bone morphogenetic protein type 2; CAV1 = caveolin 1; ENG = endoglin; HIV = human immunodeficiency virus; ICD-10-CM = *International Classification of Diseases, 10th Revision, Clinical Modification*; KCNK3 = potassium channel superfamily K member 3; PAH = pulmonary arterial hypertension; PH = pulmonary hypertension; SMAD9 = SMAD family member 9; WHO = World Health Organization. Adapted from Reference 15.

of first diagnosis of PH or PAH and/or first pharmacy claim for a PAH-specific medication, because this is especially important if PAH incidence is being studied.

We recommend that algorithms require at least one outpatient claim or at least one inpatient claim with a diagnosis of PH. A similar algorithm requiring at least two outpatient claims identified nearly the same number of patients, suggesting that at least one outpatient claim may be sufficient (16). Requiring only inpatient claims reduces sample sizes (22); however, using inpatient claims in combination with outpatient claims will increase sample size. Only FDA-approved medications for the treatment of PAH should be used. Excluding specific medications may reduce potential misclassification as PAH. First, the use of calcium channel blockers as PAH-related medication should be avoided because these are recommended for a very small number of patients with positive acute vasoreactivity testing, but they are more commonly used to treat other cardiovascular conditions (23). Second, use of the brands Viagra (Pfizer; sildenafil) and Cialis (Eli Lilly & Co.; tadalafil) on the list of PAH-specific medications should also be precluded, because these medications are indicated for erectile dysfunction (23). Alternatively, the quantity and dose dispensed can be considered to distinguish patients receiving these medications, because Revatio (Pfizer; sildenafil) and Adcirca (United Therapeutics; tadalafil) are being branded specifically for PAH treatment, and both products now have generic equivalents available. Alternatively, patients with erectile dysfunction may also be excluded to reduce this misclassification (18). The drug riociguat is approved by the FDA for two types of PH: nonoperative CTEPH (group 4 PH) and PAH (group 1 PH). If patients with groups 2 to 5 WHO classification conditions are excluded as suggested earlier, this drug should identify only patients with PAH; otherwise, patients with group 4 PH may be misclassified if they are not excluded. Requiring two PAH-specific drug classes versus only one PAH-specific drug class will likely increase PPV (i.e., reduce the number of false-positive results) (16, 17).

Two-Component Algorithms

Alternative algorithms using a combination of two components may also be useful in certain situations. Algorithms requiring diagnosis codes and PAH-specific medication but no RHC will likely produce some false-positive results. To reduce false-positive results, the following strategies could be used: *1*) require only the primary diagnosis code (ICD-9-CM code 416.0; ICD-10-CM code I27.0), although this will restrict the algorithm to patients with idiopathic PAH and exclude patients with PAH associated with other conditions; *2*) require patients to have filled PAH-specific drugs from at least two different classes; or *3*) require diagnosis of one of the PH WHO classification group 1 conditions together with the diagnosis codes; and *4*) exclude patients with diagnosis codes for PH-related conditions (e.g., chronic obstructive pulmonary disease, heart failure) listed under groups 2 to 5 of the WHO classification (19–21). Such an algorithm would be especially appropriate for studying PAH treatment patterns.

Algorithms requiring diagnosis codes and RHC will improve the true-positive cases of PAH, but sample size will be reduced. Current practice guidelines require RHC to definitively diagnose PAH; however, a notable proportion of patients within a claims-based dataset will not undergo RHC during the study period (16, 18). Requiring either RHC or echocardiography can be considered because a large proportion of patients will have undergone either one of them (11, 18). When only medical claims are available, the algorithm should require patients to have one of the WHO group 1 conditions for increasing the likelihood of truly having PAH. This algorithm may be more appropriate for PAH prevalence and assessing PAH as an outcome in WHO group 1 idiopathic PAH and associated PAH conditions. The major limitation with this approach is that the results of RHC or echocardiography are not available in claims data. Thus, just having undergone RHC or echocardiography is not an indicator of a PAH-specific diagnostic result. Another limitation is that although WHO classification group 1 conditions are most often associated with PAH, other forms of

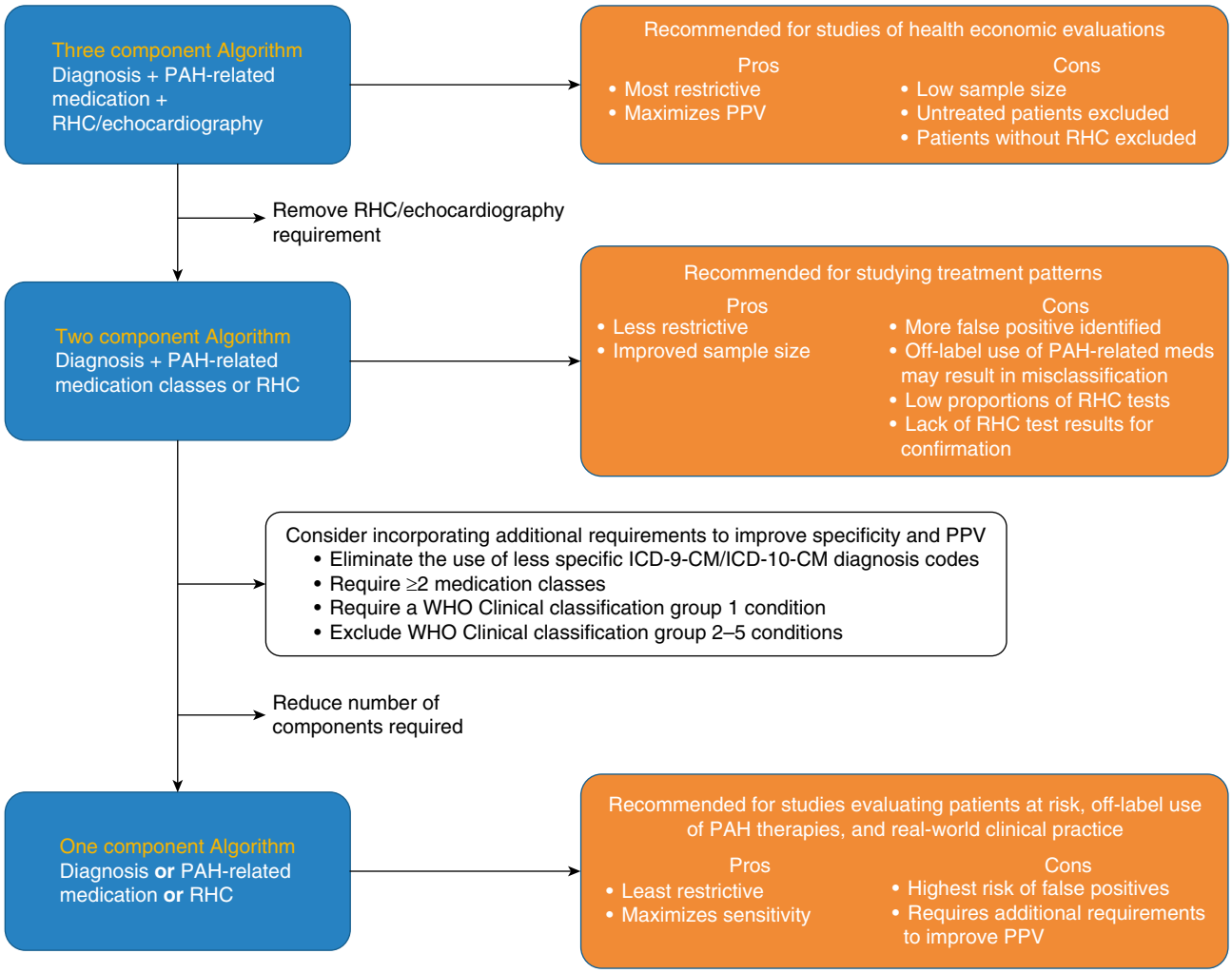

Figure 1. Algorithm variations and recommendations. ICD-9-CM = *International Classification of Diseases, Ninth Revision, Clinical Modification*; ICD-10-CM = *International Classification of Diseases, 10th Revision, Clinical Modification*; PAH = pulmonary arterial hypertension; PPV = positive predictive value; RHC = right heart catheterization; WHO = World Health Organization.

PH can also occur within these same conditions. Thus, some of the underlying WHO group 1 conditions can be associated with non–group 1 disease, including SSc (24). Patients with SSc can develop PH (traditionally described as mean pulmonary arterial pressure ⩾25 mm Hg) caused by PAH, left ventricular disease, or pulmonary fibrosis; therefore, accurate diagnosis of PAH is particularly challenging. To reduce false-positive results, requiring a claim for RHC would likely yield an improved PPV.

One-Component Algorithms

Algorithms requiring only one component are likely to identify more false-positive patients. Such an algorithm would be useful for studies evaluating patients at risk of PAH. We recommend restricting identification to ICD-9-CM code 416.0 and ICD-10-CM code I27.0 for idiopathic PAH and requiring at least two claims with these diagnosis codes when using diagnosis codes alone. Consider this algorithm when pharmacy claims are not available. Alternatively, when only pharmacy claims data are available, use PAH-specific medications (13) and require at least two different PAH-specific medication classes to improve identification of patients with PAH. Such an algorithm could be useful for studying drug use and pharmacy costs. An RHC procedure alone can be used when only patient charts or EMR data are available, given the availability of procedure results in these datasets.

Limitations of Algorithms

There are some common limitations inherent to claims data. Claims lack information on disease severity, which can be useful for the evaluation of patients with PAH. As described above, coding with ICD-9-CM and ICD-10-CM codes can result in errors or can be influenced by reimbursement decisions. Patients may not be continuously enrolled in the health insurance plan or may die during the study period, resulting in exclusion from the study or loss to follow-up, which can

adversely affect the sample size and bias results if these exclusions are systematically associated with disease or exposure status.

Claims-based analyses in PH have additional potential limitations, including the lack of specificity of ICD-9-CM and ICD-10-CM codes in relation to WHO clinical classification. The use of claims data is further complicated by the lack of laboratory test results. Although claims provide insight into whether RHC was conducted for diagnostic purposes, the results of that test are not available in the claims data, and thus perfect accuracy with respect to PAH diagnosis (or WHO classification) is not possible. For example, patients who do not meet hemodynamic criteria for PAH would meet algorithm requirements for an RHC claim but would be misclassified as having PAH. For most of these algorithms to be successful, both medical and pharmacy claims data must be available. In the absence of pharmacy claims data, procedure codes for RHC or echocardiography could be used in combination with PH-related diagnosis codes. In addition, some of the PAH-specific medications may be used for off-label conditions, which may confound or bias the results. Although ICD-10-CM codes provide more specification pertaining to PAH, they are not solely sufficient to accurately identify patients with PAH, and this may, in fact, make future claims-based analyses more complicated because researchers must determine how these codes are clinically used.

There is also significant clinical practice variation in the diagnosis and treatment of PH and PAH, which may or may not adhere to consensus guidelines; this could result in misclassification bias and affect the conclusions drawn from using administrative claims data. The results from these studies may be hypothesis generating but should be used cautiously.

Next Steps in PAH-related Real-World Evidence Research

Identifying patients with PAH using administrative claims data can be challenging. Depending on the research question and availability of the data, several algorithms could be used in claims-based research involving patients with PAH. Correctly identifying patients with PAH in claims-based studies can improve the value, credibility, and accuracy of research findings and the use of real-world evidence in support of FDA regulatory decisions, and it can more precisely inform formulary decision making. Algorithms requiring an FDA-approved medication for PAH are recommended. Requiring a claim for a PAH-related medication will capture a notable proportion of patients with PAH, reduce false-positive results, and likely improve the overall performance of the algorithm when used in combination with a PH diagnosis. Although this report provides algorithms with anticipated results, primary data were not used to test the performance characteristics of the algorithms. It will be prudent for future researchers to validate the algorithms and to assess their performance against direct identification of patients with PAH through review of primary clinical and diagnostic data. Such studies would add valuable information about the performance characteristics of our proposed algorithms. ■

Author disclosures are available with the text of this article at www.atsjournals.org.

References

1 Simonneau G, Montani D, Celermajer DS, Denton CP, Gatzoulis MA, Krowka M, *et al*. Haemodynamic definitions and updated clinical classification of pulmonary hypertension. *Eur Respir J* 2019;53: 1801913.

2 Simonneau G, Gatzoulis MA, Adatia I, Celermajer D, Denton C, Ghofrani A, *et al*. Updated clinical classification of pulmonary hypertension. *J Am Coll Cardiol* 2013;62(25, Suppl):D34–D41.

3 Galiè N, Humbert M, Vachiery JL, Gibbs S, Lang I, Torbicki A, *et al*.; ESC Scientific Document Group. 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 Heart J* 2016;37: 67–119.

4 Humbert M, Montani D, Evgenov OV, Simonneau G. Definition and classification of pulmonary hypertension. *Handb Exp Pharmacol* 2013;218:3–29.

5 Schraufnagel DE, editor. Breathing in America: diseases, progress, and hope [Internet]. New York: American Thoracic Society; 2010 [accessed 2019 Mar 1]. Available from: https://www.thoracic.org/patients/patient-resources/breathing-in-america/resources/breathing-in-america.pdf.

6 Hoeper MM, Bogaard HJ, Condliffe R, Frantz R, Khanna D, Kurzyna M, *et al*. Definitions and diagnosis of pulmonary hypertension. *J Am Coll Cardiol* 2013;62(25, Suppl):D42–D50.

7 Cadarette SM, Wong L. An introduction to health care administrative data. *Can J Hosp Pharm* 2015;68:232–237.

8 Choudhry NK. Randomized, controlled trials in health insurance systems. *N Engl J Med* 2017;377:957–964.

9 Jarow JP, LaVange L, Woodcock J. Multidimensional evidence generation and FDA regulatory decision making: defining and using "real-world" data. *JAMA* 2017;318:703–704.

10 Aronson JK. Rare diseases and orphan drugs. *Br J Clin Pharmacol* 2006; 61:243–245.

11 Fischer A, Kong AM, Swigris JJ, Cole AL, Raimundo K. All-cause healthcare costs and mortality in patients with systemic sclerosis with lung involvement. *J Rheumatol* 2018;45:235–241.

12 Papani R, Duarte AG, Lin YL, Kuo YF, Sharma G. Pulmonary arterial hypertension associated with interferon therapy: a population-based study. *Multidiscip Respir Med* 2017;12:1.

13 Waxman A, Chen SY, Boulanger L, Watson JA, Golden G. Factors associated with adherence to phosphodiesterase type 5 inhibitors for the treatment of pulmonary arterial hypertension. *J Med Econ* 2013; 16:298–306.

14 Choi YM, Famenini S, Wu JJ. Incidence of pulmonary arterial hypertension in patients with psoriasis: a retrospective cohort study. *Perm J* 2017;21:16-073.

15 Mathai SC, Mathew S. Breathing (and coding?) a bit easier: changes to International Classification of Disease coding for pulmonary hypertension. *Chest* 2018;154:207–218.

16 Burger CD, Pruett JA, Lickert CA, Berger A, Murphy B, Drake W III. Prostacyclin use among patients with pulmonary arterial hypertension in the United States: a retrospective analysis of a large health care claims database. *J Manag Care Spec Pharm* 2018;24:291–302.

17 Papani R, Sharma G, Agarwal A, Callahan SJ, Chan WJ, Kuo YF, *et al*. Validation of claims-based algorithms for pulmonary arterial hypertension. *Pulm Circ* 2018;8:2045894018759246.

18 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.
19 Kirson NY, Birnbaum HG, Ivanova JI, Waldman T, Joish V, Williamson T. Excess costs associated with patients with pulmonary arterial hypertension in a US privately insured population. *Appl Health Econ Health Policy* 2011;9:293–303.
20 Kirson NY, Birnbaum HG, Ivanova JI, Waldman T, Joish V, Williamson T. Prevalence of pulmonary arterial hypertension and chronic thromboembolic pulmonary hypertension in the United States. *Curr Med Res Opin* 2011;27:1763–1768.
21 Said Q, Martin BC, Joish VN, Kreilick C, Mathai SC. The cost to managed care of managing pulmonary hypertension. *J Med Econ* 2012;15: 500–508.
22 Burger CD, Ozbay AB, Lazarus HM, Riehle E, Montejano LB, Lenhart G, *et al*. Treatment patterns and associated health care costs before and after treatment initiation among pulmonary arterial hypertension patients in the United States. *J Manag Care Spec Pharm* 2018;24: 834–842.
23 Copher R, Cerulli A, Watkins A, Laura Monsalvo M. 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.
24 Launay D, Sobanski V, Hachulla E, Humbert M. Pulmonary hypertension in systemic sclerosis: different phenotypes. *Eur Respir Rev* 2017;26: 170056.
25 Yang X, Sanders KN, Mardekian J, Mychaskiw MA, Thomas J III. Associations between sildenafil use and changes in days of hospitalization in a population with pulmonary arterial hypertension associated with connective tissue disease. *Clin Ther* 2015;37: 1055–1063.
26 Angalakuditi M, Edgell E, Beardsworth A, Buysman E, Bancroft T. Treatment patterns and resource utilization and costs among patients with pulmonary arterial hypertension in the United States. *J Med Econ* 2010;13:393–402.
27 Berger A, Edelsberg J, Teal S, Mychaskiw MA, Oster G. Changes in healthcare utilization and costs associated with sildenafil therapy for pulmonary arterial hypertension: a retrospective cohort study. *BMC Pulm Med* 2012;12:75.
28 Sikirica M, Iorga SR, Bancroft T, Potash J. The economic burden of pulmonary arterial hypertension (PAH) in the US on payers and patients. *BMC Health Serv Res* 2014;14:676.
29 Burke JP, Hunsche E, Régulier E, Nagao M, Buzinec P, Drake Iii W. 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(3, Suppl):s47–s58.
30 Dufour R, Pruett J, Hu N, Lickert C, Stemkowski S, Tsang Y, *et al*. Healthcare resource utilization and costs for patients with pulmonary arterial hypertension: real-world documentation of functional class. *J Med Econ* 2017;20:1178–1186.