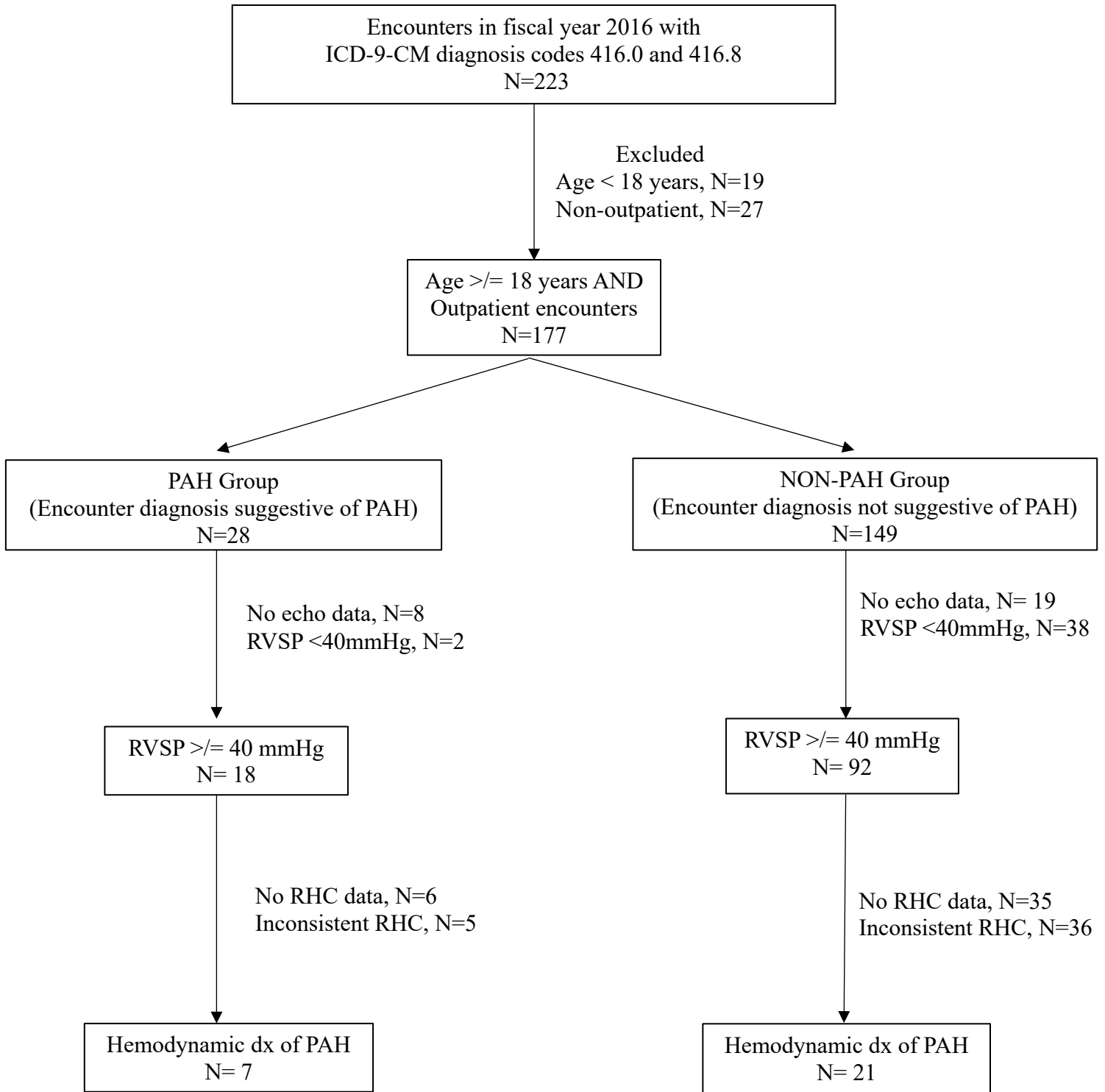


**SUPPLEMENTARY MATERIAL**



**FIGURE 1e** Validation 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 1e** Baseline Characteristics of the validation cohort: patients in pulmonary arterial hypertension (PAH) and non-PAH groups.

	PAH (n=28)	Non-PAH (n=149)	P-value
Age: Mean (SD)	66.39 (24.87)	67.28 (15.14)	0.8561
<30 years	5 (17.86%)	3 (2.01%)	
31-40 years	0 (0.00%)	8 (5.37%)	
41-50 years	3 (10.71%)	6 (4.03%)	
51-60 years	1 (3.57%)	23 (15.44%)	
61-70 years	3 (10.71%)	41 (27.52%)	
71-80 years	7 (25.00%)	36 (24.16%)	
81-90 years	5 (17.86%)	29 (19.46%)	
90+ years	4 (14.29%)	3 (2.01%)	
Sex			0.2154
Female	15 (48.39%)	93 (60.00%)	
Male	16 (51.61%)	62 (40.00%)	
Race			1.000
Not Hispanic or Latino	28 (100%)	147 (98.66%)	
Unknown	0 (0.00%)	0 (0.00%)	
Hispanic or Latino	0 (0.00%)	2 (1.34%)	
Comorbidities			
Hypertension	19 (67.86%)	111 (74.50%)	0.4654
Congestive Heart Failure	7 (25.00%)	64 (42.95%)	0.0210
Sleep Disordered Breathing	6 (21.43%)	46 (30.87%)	0.3141
Diabetes Mellitus	8 (28.57%)	55 (36.91%)	0.3977
Chronic Pulmonary Disease	5 (17.86%)	31 (20.81%)	0.7221
Atrial Fibrillation	5 (17.86%)	34 (22.82%)	0.5611
Obesity	10 (35.71%)	69 (46.31%)	0.3008
Coronary Artery Disease	11 (39.29%)	40 (26.85%)	0.1823
Valvular Heart Disease	6 (21.43%)	21 (14.09%)	0.3881
Connective Tissue Disorder	2 (7.14%)	19 (12.75%)	0.5356
Liver Disease	4 (14.29%)	9 (6.04%)	0.1291
Atrial Flutter	0 (0.00%)	4 (2.68%)	1.0000
Congenital Heart Disease	7 (25.00%)	13 (8.72%)	0.0210
HIV	0 (0.00%)	1 (0.67%)	1.0000
Interstitial Lung Disease	2 (7.14%)	19 (12.75%)	0.5356

## **2x2 ANALYSIS FOR CLAIMS ALGORITHMS**

(Dx- diagnosis; EMR- Electronic medical record; RHC- Right heart catheterization; PAHRx- PAH-specific therapies; PPV- positive predictive value; NPV- negative predictive value).

### *DEVELOPMENT ALGORITHMS*

ICD-9-CM codes 416.0 and 416.8	Gold Standard +	Gold Standard -	Total
Test +	49+3(4.92)=63.76	619.24	683
Test -	--	--	--
Total	--	--	--

#### **A) Claims-based algorithms**

PPV =  $63.76/683 = 9.33\%$

ICD codes + At least one PAHRx	Gold Standard +	Gold Standard -	Total
Test +	43	22+12(4.92)=81.04	124.04
Test -	6+3(4.92)=20.76	120+85(4.92)=538.2	558.96
Total	63.76	619.24	683

Sensitivity =  $43/63.46 = 67.44\%$

Specificity =  $538.2/619.24 = 86.91\%$

PPV =  $43/124.04 = 34.67\%$

NPV =  $538.2/558.96 = 96.29\%$

ICD codes + Two or more classes PAHRx	Gold Standard +	Gold Standard -	Total
Test +	18	4+1(4.92)=8.92	22+1(4.92)=26.92
Test -	31+3(4.92)=45.76	138+96(4.92)=610.32	169+99(4.92)=656.08
Total	63.76	619.24	683

Sensitivity =  $18/63.76 = 28.23\%$

Specificity =  $610.32/619.24 = 98.56\%$

PPV =  $18/26.92 = 66.86\%$

NPV =  $610.32/656.08 = 93.03\%$

#### **B) EMR Based algorithms**

ICD9 + EMR encounter dx	Gold Standard +	Gold Standard -	Total
Test +	49	142	191
Test -	3(4.92)=14.76	97(4.92)=477.24	492
Total	63.76	619.24	683

Sensitivity =  $49/63.76 = 76.85\%$

Specificity =  $477.24/619.24 = 77.07\%$

PPV =  $49/191 = 25.65\%$

NPV =  $477.24/492 = 97.00\%$

ICD codes + EMR encounter dx + echo	Gold Standard +	Gold Standard -	Total
-------------------------------------	-----------------	-----------------	-------

Test +	49	135	184
Test -	3(4.92)=14.76	97(4.92) + 7= 484.24	499
Total	63.76	619.24	683

Sensitivity =  $49/63.76=76.85\%$

Specificity =  $484.24/619.24=78.20\%$

PPV =  $49/184=26.63\%$

NPV =  $484.24/499=97.04\%$

ICD codes + EMR encounter dx + echo + RHC	Gold Standard +	Gold Standard -	Total
Test +	49	53	102
Test -	3(4.92)=14.76	97(4.92) + 89 = 566.24	581
Total	63.76	619.24	683

Sensitivity =  $49/63.76=76.85\%$

Specificity =  $566.24/619.24=91.44\%$

PPV =  $49/102=48.04\%$

NPV =  $566.24/581=97.46\%$

ICD codes + EMR encounter dx + echo + RHC + at least one PAHRx	Gold Standard +	Gold Standard -	Total
Test +	43	19	62
Test -	6+3(4.92)=20.76	123+97(4.92)=600.24	621
Total	63.76	619.24	683

Sensitivity =  $43/63.76 = 67.44\%$

Specificity =  $600.24/619.24 = 96.93\%$

PPV =  $43/62 = 69.35\%$

NPV =  $600.24/621 = 96.66\%$

ICD codes + EMR encounter dx + at least one PAHRx	Gold Standard +	Gold Standard -	Total
Test +	43	22	65
Test -	6+ 3*4.92 =20.76	120 + 97*4.92 =597.24	618
Total	63.76	619.24	683

PPV =  $43/65 = 66.15\%$

NPV =  $597.24/618 = 96.64\%$

Sensitivity =  $43/63.76 = 67.44\%$

Specificity =  $597.24/619.24 = 96.45\%$

VALIDATION ALGORITHMS

**A) Claims-based algorithms**

ICD-9-CM codes 416.0 and 416.8	Gold Standard +	Gold Standard -	Total
Test +	28	149	177
Test -	--	--	--
Total	--	--	--

PPV =  $28/177 = 15.82\%$

ICD codes + At least one PAHRx	Gold Standard +	Gold Standard -	Total
Test +	18	27	45
Test -	10	122	132
Total	28	149	177

PPV =  $18/45 = 40.00\%$

NPV =  $122/132 = 92.42\%$

Sensitivity =  $18/28 = 64.29\%$

Specificity =  $122/149 = 81.88\%$

ICD codes + More than one PAHRx	Gold Standard +	Gold Standard -	Total
Test +	12	9	21
Test -	16	140	156
Total	28	149	177

PPV =  $12/21 = 57.14\%$

NPV =  $139/156 = 89.74\%$

Sensitivity =  $12/28 = 42.86\%$

Specificity =  $140/149 = 93.96\%$

**B) EMR Based algorithms**

ICD codes + Encounter dx	Gold Standard +	Gold Standard -	Total
Test +	7	21	28
Test -	21	128	149
Total	28	149	177

PPV =  $7/28 = 25.00\%$

NPV =  $128/149 = 85.91\%$

Sensitivity =  $7/28 = 25.00\%$

Specificity =  $128/149 = 85.91\%$

ICD codes + Encounter dx + echo	Gold Standard +	Gold Standard -	Total
Test +	7	13	20
Test -	21	136	157

Total	28	149	177
-------	----	-----	-----

PPV =  $7/20=35.00\%$

NPV =  $136/157=86.62\%$

Sensitivity =  $7/28=25.00\%$

Specificity =  $136/149=91.28\%$

ICD codes + Encounter dx + echo + RHC	Gold Standard +	Gold Standard -	Total
Test +	7	5	12
Test -	21	144	165
Total	28	149	177

PPV =  $7/12=58.33\%$

NPV =  $144/165=87.27\%$

Sensitivity =  $7/28=25.00\%$

Specificity =  $144/149=96.64\%$

ICD codes + Encounter dx + echo + RHC + at least one PAHRx	Gold Standard +	Gold Standard -	Total
Test +	5	3	8
Test -	23	146	169
Total	28	149	177

PPV =  $5/8= 62.50\%$

NPV =  $146/169= 86.39\%$

Sensitivity =  $5/28= 17.86\%$

Specificity =  $146/149= 97.99\%$

ICD9 + ENC DX + at least one PAHRx	Gold Standard +	Gold Standard -	Total
Test +	5	6	11
Test -	23	143	166
Total	28	149	177

PPV =  $5/11= 45.45\%$

NPV =  $143/166= 86.14\%$

Sensitivity =  $5/28= 17.86\%$

Specificity =  $143/149= 95.97\%$