February 27th, 2024
Community call update
Week 1: Alzheimer's disease (AD)
- AD Literature review complete
- AD definition replicated
- Community call update
- Phenotype Representation analysis for AD
- Cohort Diagnostics PheValuator run
- Atlas demo Cohort Diagnostics review

Week 2: Non-Small Cell and Small Cell Lung Cancer (NSCLC)
- LC Literature review complete
- Community call update
- NSCLC definition replication
- A discussion on what can be replicated

Week 3: Major Depressive Disorder (MDD)
- MDD Literature review complete
- Community call update
- MDD definition replication
- PAH data extr. in progress
- Atlas QA Cohort Diagnostics review
- MDD Literature review complete

Week 4: Pulmonary Arterial Hypertension (PAH)
- PAH Literature review complete
- Community call update
- PAH definition replication
- Atlas QA Cohort Diagnostics review
Pulmonary Hypertension & Pulmonary Arterial Hypertension

Pulmonary Hypertension: Mean Pulmonary Artery Pressure > 20mmHg (normal 8-20)

https://en.wikipedia.org/wiki/Capillary

https://forums.ohdsi.org/t/putting-the-phenotype-phebraury-day-20/15983
Pulmonary Arterial Hypertension: Diagnosis

- Group 1 – Pulmonary Arterial Hypertension
  - Primary / idiopathic
  - Secondary to CTDs, drugs, toxins, infections.

- Group 2 – PH Due to Left Heart Disease

- Group 3 – PH due to Chronic Lung Disease and/or hypoxemia

- Group 4 – PH due to Pulmonary arterial obstructions
  - Chronic Thromboembolic Pulmonary Hypertension, CTEPH

- Group 5 – Unclear, multifactorial mechanisms

Precapillary: Blue
(mostly) Post capillary: Red
Mixed: Purple

<table>
<thead>
<tr>
<th>ICD-9 Code</th>
<th>ICD-10 Code</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>416.0</td>
<td>I27.0</td>
<td>Primary pulmonary hypertension</td>
</tr>
<tr>
<td>416.2</td>
<td>I27.2</td>
<td>Other secondary pulmonary hypertension</td>
</tr>
<tr>
<td>416.8</td>
<td></td>
<td>Other chronic pulmonary heart diseases</td>
</tr>
<tr>
<td>416.9</td>
<td></td>
<td>Chronic pulmonary heart disease, unspecified</td>
</tr>
</tbody>
</table>

ICD-9 and ICD-10 Coding systems do not reflect the Clinical Classification of PH

PVRI RWE Working Group; Jan 31 2024; with permission
Pulmonary Arterial Hypertension: Diagnosis

**Pulmonary Hypertension**: Mean PAP > 20mmHg

**Pulmonary Arterial Hypertension**: Mean PAP > 20mmHg & wedge Pressure < 15 mmHg & PVR > 2 Wood Units

https://en.wikipedia.org/wiki/Cardiac_catheterization
Pulmonary Arterial Hypertension: Therapy

- PAH patients have:
  - Elevated levels of Endothelin-1 (vasoconstrictor)
    - Endothelin Receptor Antagonists
  - Low levels of endogenous Nitrous Oxide (NO)
    - PDE-5i; soluble cGMP stimulators
  - Low levels of Prostacyclin
    - Prostacyclin Derivatives

- Calcium Channel Blockers in NO responsive
- Anticoagulation (controversial)
- Some may be tried in other PH groups (possible elements of PAH in Group 2-5)
- Dual therapy now used in most PAH patients if diagnosed when symptomatic

https://www.sciencedirect.com/science/article/pii/S0378517322003477?via%3Dihub (Creative Commons)
What did we do?

40 manuscripts identified for review (2020 - 2024)

15 manuscripts filtered based on automated scoring mechanism

Data extraction performed on 11 manuscripts - mainly from the USA

• Prospective studies
• Registry studies
• Additional papers added from a review article

Thank you:
Bolu Oluwalade, Thamir Alshammari, Septi Melisa, Andreas Weinberger Rosen, Mengchun Gong, Milou Brand, Judy Racoosin
Glance at phenotype definitions

Common patterns in concept sets:
- I27.0 (Primary pulmonary hypertension) only or add
- I27.20, I27.21 (Pulmonary hypertension, unspecified, Secondary pulmonary arterial hypertension)
- and I27.89 (Other specified pulmonary heart

Common patterns in phenotype definitions:
- One or two of the diagnoses
- commonly require a treatment, with variation in the treatment list
- exclusion of differential diagnoses (chronic thromboembolic pulmonary hypertension)

More details when we replicate the cohorts!
Clinical evaluation of code-based algorithms to identify patients with pulmonary arterial hypertension in healthcare databases

Eva-Maria Didden, Di Lu, Andrew Hsi, Monika Brand, Haley Hedlin, Roham T. Zamanian

First published: 08 February 2024 | https://doi.org/10.1002/pul2.12333

Abstract

Identifying Patients with Pulmonary Arterial Hypertension Using Administrative Claims Algorithms - PubMed (nih.gov)
PAH Phenotype Validation

OHDSI Community Call
February 27th, 2024

Eva-Maria Didden

On behalf of the study teams:

*PheValuator-based PAH phenotype validation:* Viviane Sprecher, EMD, Joel Swerdel, Audrey Muller

*Clinical PAH phenotype validation through database linkage:* EMD, Di Lu, Andrew Hsi, Monika Brand, Haley Hedlin, Roham T. Zamanian
Background – PAH phenotype validation

Pulmonary Arterial Hypertension [PAH]:
• Subgroup of Pulmonary Hypertension [PH] with diverse etiologies.
• Rare and life-threatening, but treatable (not curable!).
• Unspecific symptoms, high misdiagnosis rates, delayed diagnosis.

The challenge: identifying PAH patients in observational healthcare databases:
• PAH diagnosis codes:
  ▪ release of P(A)H-specific ICD codes only in Oct ‘17.
  ▪ might represent a rule-out diagnosis or suspicion of the disease (i.e., PAH code used for specialist referral or PAH screening purposes).
• PAH drug codes: might be used off-label for treatment of other forms of PH.

Common solution: Well-defined temporal sequences of diagnosis, procedure, drug, and/or exclusionary codes ➔ PAH phenotype algorithms.
Objectives – PAH phenotype validation

Most recent publication*:
To demonstrate PAH phenotype validation through linkage of an EHR database with a PH-specific clinical database.

Previous PheValuator work**:
To validate PAH phenotype algorithms identified via a systematic literature search in US health insurance claims databases, using PheValuator.


**Sprecher VP, Didden EM, Swerdel JN, Muller A. Evaluation of code-based algorithms to identify pulmonary arterial hypertension and chronic thromboembolic pulmonary hypertension patients in large administrative databases. Pulm Circ 2020;10:2045894020961713
Methods – clinical PAH phenotype validation

Databases, linked through unique patient identifiers:
• Stanford Healthcare administrative EHR database to apply the algorithms.
• Stanford Vera Moulton Wall Center (VMWC) clinical PH database to perform clinical case validation and assessment of algorithm performance.

PAH phenotype algorithms for validation:
• Six published algorithms.
• Ten additional clinically meaningful algorithms.

Algorithm performance metrics:
• True Positives, True Negatives, False Positives, False Negatives.
• Sensitivity, Specificity, Positive Predictive Value (PPV), Negative Predictive Value (NPV).
Eligible PH patients:

- ≥ 18 years of age,
- linkable records,
- ≥1 Right Heart Catheterization (RHC) at Stanford VMWC,
- a confirmed PH diagnosis (mean pulmonary arterial pressure ≥25 mmHg),
- ≥6 months of observation after their first RHC at Stanford VMWC (unless they died),
- in case of an early death, ≥1 visit between the first RHC and death.

*Patients with a PAH diagnosis and no PH WHO Group II–V diagnosis at any point in their medical history.
†Patients with a PAH diagnosis and a PH WHO Group II–V diagnosis at any point in their medical history.

For sensitivity analysis
### Results:

<table>
<thead>
<tr>
<th>Algorithm ID</th>
<th>Algorithm includes</th>
<th>Sensitivity (95% CI)</th>
<th>Specificity (95% CI)</th>
<th>PPV (95% CI)</th>
<th>NPV (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Published algorithms</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 / 28678692</td>
<td>Dx, RHC/TTE, Rx, Excl</td>
<td>1.000 (0.993, 1.000)</td>
<td>0.000 (0.000, 0.000)</td>
<td>0.775 (0.742, 0.805)</td>
<td>0.000 (0.000, 0.000)</td>
</tr>
<tr>
<td>1 / 27851838</td>
<td>Dx, RHC/TTE, Rx</td>
<td>0.978 (0.962, 0.988)</td>
<td>0.117 (0.072, 0.177)</td>
<td>0.792 (0.760, 0.822)</td>
<td>0.613 (0.421, 0.781)</td>
</tr>
<tr>
<td>6 / 30421652</td>
<td>Dx, RHC/TTE, Rx, Excl</td>
<td>0.953 (0.932, 0.969)</td>
<td>0.383 (0.307, 0.462)</td>
<td>0.841 (0.810, 0.869)</td>
<td>0.704 (0.597, 0.797)</td>
</tr>
<tr>
<td>3 / 28762848</td>
<td>Dx, RHC/TTE, Rx, Excl</td>
<td>0.953 (0.932, 0.969)</td>
<td>0.383 (0.307, 0.462)</td>
<td>0.841 (0.810, 0.869)</td>
<td>0.704 (0.597, 0.797)</td>
</tr>
<tr>
<td>5 / 30566510</td>
<td>Rx, Excl</td>
<td>0.068 (0.048, 0.092)</td>
<td>1.000 (0.977, 1.000)</td>
<td>1.000 (0.907, 1.000)</td>
<td>0.237 (0.206, 0.271)</td>
</tr>
<tr>
<td>4 / 29485908</td>
<td>Dx, RHC/TTE, Rx</td>
<td>0.041 (0.026, 0.061)</td>
<td>1.000 (0.977, 1.000)</td>
<td>1.000 (0.851, 1.000)</td>
<td>0.232 (0.201, 0.265)</td>
</tr>
<tr>
<td><strong>Additional (unpublished) algorithms</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9 / NA</td>
<td>Dx, RHC/TTE, Rx, Excl</td>
<td>0.998 (0.990, 1.000)</td>
<td>0.000 (0.000, 0.000)</td>
<td>0.775 (0.742, 0.804)</td>
<td>0.000 (0.000, 0.000)</td>
</tr>
<tr>
<td>7 / NA</td>
<td>Dx, RHC/TTE, Rx</td>
<td>0.996 (0.987, 0.999)</td>
<td>0.018 (0.003, 0.053)</td>
<td>0.777 (0.745, 0.807)</td>
<td>0.600 (0.146, 0.947)</td>
</tr>
<tr>
<td>8 / NA</td>
<td>Dx, RHC/TTE, Rx</td>
<td>0.996 (0.987, 0.999)</td>
<td>0.018 (0.003, 0.053)</td>
<td>0.777 (0.745, 0.807)</td>
<td>0.600 (0.146, 0.947)</td>
</tr>
<tr>
<td>10 / NA</td>
<td>Dx, RHC/TTE, Rx, Excl</td>
<td>0.944 (0.922, 0.961)</td>
<td>0.383 (0.307, 0.462)</td>
<td>0.840 (0.809, 0.868)</td>
<td>0.666 (0.561, 0.761)</td>
</tr>
<tr>
<td>11 / NA</td>
<td>Dx, RHC/TTE, Rx, Excl</td>
<td>0.845 (0.813, 0.874)</td>
<td>0.562 (0.481, 0.639)</td>
<td>0.869 (0.837, 0.896)</td>
<td>0.514 (0.438, 0.589)</td>
</tr>
<tr>
<td>12 / NA</td>
<td>Dx, RHC/TTE, Rx, Excl</td>
<td>0.774 (0.737, 0.808)</td>
<td>0.488 (0.408, 0.567)</td>
<td>0.838 (0.804, 0.869)</td>
<td>0.385 (0.318, 0.455)</td>
</tr>
<tr>
<td>13 / NA</td>
<td>Dx, RHC/TTE, Rx</td>
<td>0.509 (0.466, 0.551)</td>
<td>0.617 (0.537, 0.692)</td>
<td>0.821 (0.776, 0.859)</td>
<td>0.267 (0.223, 0.315)</td>
</tr>
<tr>
<td>13b / NA</td>
<td>Dx, RHC/TTE, Rx</td>
<td>0.509 (0.466, 0.551)</td>
<td>0.617 (0.537, 0.692)</td>
<td>0.821 (0.776, 0.859)</td>
<td>0.267 (0.223, 0.315)</td>
</tr>
<tr>
<td>7b / NA</td>
<td>Dx, RHC/TTE, Rx, Excl</td>
<td>0.069 (0.050, 0.094)</td>
<td>1.000 (0.977, 1.000)</td>
<td>1.000 (0.907, 1.000)</td>
<td>0.237 (0.206, 0.271)</td>
</tr>
<tr>
<td>9b / NA</td>
<td>Dx, RHC/TTE, Rx</td>
<td>0.068 (0.048, 0.092)</td>
<td>1.000 (0.977, 1.000)</td>
<td>1.000 (0.907, 1.000)</td>
<td>0.237 (0.206, 0.271)</td>
</tr>
</tbody>
</table>

**Abbreviations:**
- **Dx:** diagnostic code;
- **RHC:** right heart catheterization;
- **TTE:** transthoracic echocardiography;
- **Rx:** pharmacy claim;
- **Excl:** exclusionary codes;
- **temporal component:**
- **PMID:** PubMed ID;
- **CI:** confidence interval.
Sensitivity-Specificity Trade-Off

Diagnostics
- True positives
- False positives
- False negatives
- True negatives

Algorithm ID
- 8
- 1
- 3
- 11
- 13b
- 5

Dx
RHC/TTE
Rx
Excl

Percent
- 77%
- 76%
- 74%
- 66%
- 39%
- 72%

Algorithm ID
- 22%
- 20%
- 14%
- 12%
- 38%
- 23%
Summary

There is no “best” algorithm:

- Inclusive algorithms with high sensitivity (> 0.94) are non-specific (specificity <0.40).
- Selective algorithms with high specificity (1.00) are not sensitive (sensitivity <0.10).
- Algorithms with a reasonable balance of sensitivity and specificity (both >0.50) typically consist of well-defined temporal sequences of procedure, diagnosis, and drug codes.
- In line with expert findings and recommendations for PAH algorithm development*.

Notes from additional/sensitivity analyses:

- Across all algorithms, only minor random variations in characteristics of correctly identified PAH patients.
- Same findings when excluding patients with both a PAH and a PH WHO Group II–V diagnosis from study.

Recommendations

1. Tailor algorithm selection/design to the specific research question.  
   ➡️ Is a sensitive, specific, or balanced algorithm required?

2. Revisit research question and assess all relevant patient characteristics.  
   ➡️ Should additional selection criteria/codes be included in the algorithm?

3. Include temporal components in the algorithm, as appropriate.  
   ➡️ What should be the temporal sequence of events/codes? This can vary between regions and healthcare systems.

4. Describe your algorithm(s) in detail in your publication.
Strengths and Limitations

• **This study provides a robust case validation:** all true P(A)H patients could be classified based on RHC – the gold standard - in the clinical database.

• Stanford is a center of specialized PH care ➔ PAH prevalence may be biased among PH patients ➔ generalizability of results may be impacted.

• The team suggests performing additional case validation across databases and healthcare settings based on the presented findings.
## Comparison with PheValuator study

<table>
<thead>
<tr>
<th>Databases</th>
<th>Clinical validation study</th>
<th>PheValuator study</th>
</tr>
</thead>
<tbody>
<tr>
<td>EHRs linked to a clinical PH database</td>
<td>3 US claims databases (general population)</td>
<td></td>
</tr>
<tr>
<td>PAH prevalence in database</td>
<td>78% in clinical PH database</td>
<td>0.16%–0.87%,</td>
</tr>
<tr>
<td>Ground truth</td>
<td>RHC (gold standard diagnostic test)</td>
<td>Predictions by PheValuator mathematical models that estimated the probability of each patient having PAH</td>
</tr>
<tr>
<td>Data available</td>
<td>In- and outpatient information were available in the Stanford EHR database but could not be distinguished from each other in-/outpatient algorithm components could not be considered</td>
<td>In- and outpatient information could be distinguished from each other.</td>
</tr>
</tbody>
</table>

### Findings

Algorithm rankings by sensitivity, specificity, PPN, and NPV and overall conclusions were largely similar.

---

EHR, electronic health records; PAH, pulmonary arterial hypertension; RHC, right heart catheterisation.
Both phenotype validation studies contribute to:

- having a range of universally accepted fit-for-purpose PAH phenotype algorithms, tailored to address different types of research questions.

- informing future phenotype validation work in coded healthcare databases, especially in rare or complex diseases.
References:

- Sprecher VP, Didden EM, Swerdel JN, Muller A, *Evaluation of code-based algorithms to identify pulmonary arterial hypertension and chronic thromboembolic pulmonary hypertension patients in large administrative databases*, Pulm Circ 2020;10:2045894020961713

Published algorithms validated using PheValuator and via database linkage:
Next steps

• PAH cohort replication (sign up in the sheet)
• MDD cohort diagnostics review
• PAH cohort diagnostics review next Monday
• Study package (Jamie in contact with data partners)
• Open call to plan the manuscript