EE68

Direct Medical Costs Saved by Early Pulmonary Arterial Hypertension Detection Using a Machine Learning Algorithm

Tanvee Thakur[1]; Neto Coulibaly[1]*; Praveen Kumar Marimuthu[2]; Purushotham Krishnappa[2]; Sanjeev Nadapurohit[2]; Jason Maron[2]; Anna Watzker[1]; Anna Hemnes[3]; Hilary Dubrock[4]

[1]Merck & Co., Inc., Rahway, NJ, USA; [2]Nference, Inc., Cambridge, MA, USA; [3]Vanderbilt University Medical Center, Nashville, TN, USA; [4]Mayo Clinic, Rochester, MN, USA

*Presenter of the study.

Background

- Pulmonary arterial hypertension (PAH) is a rare disease, often resulting in delayed diagnosis due to nonspecific symptoms at disease onset[1]
- Delayed diagnosis is associated with more advanced disease at presentation, worse clinical outcomes, and increased health care resource utilization (HCRU), including more frequent hospitalizations, diagnostic testing, and specialist visits[1]
- This study estimates the prediagnosis direct medical costs (DMCs) that could be saved through earlier PAH detection using a machine learning (ML) algorithm

Methods

Study design

- Retrospective study of adult patients from HIPAA-compliant electronic health record (EHR) Mayo Clinic dataset (2015-2024)
- The index date was defined as the earliest recorded date of diagnosis by either International Classification of Diseases (ICD) code or natural language processing (NLP) mention

Figure 1. CONSORT diagram for the PAH study cohort

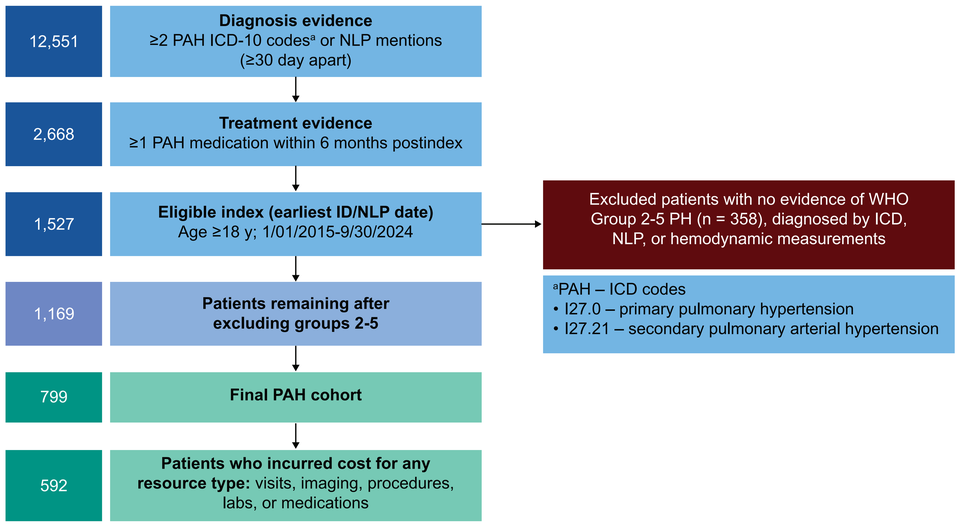
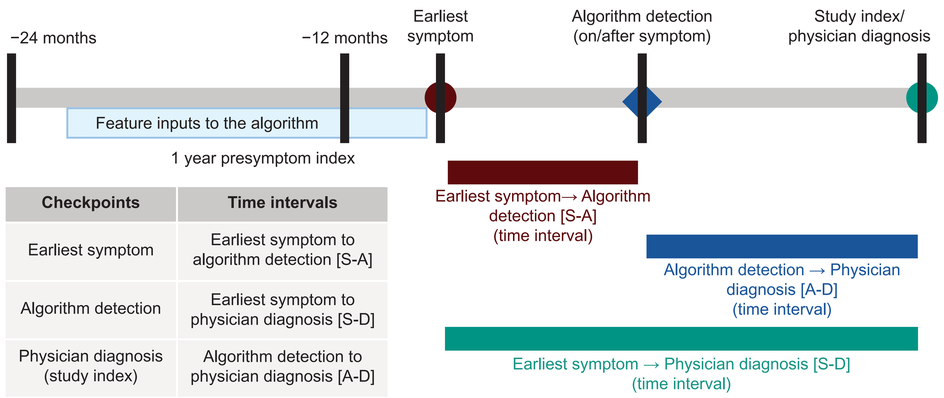


Figure 2. Key checkpoints and time intervals during a patient's care journey



Checkpoints	Time intervals
Earliest symptom	Earliest symptom to algorithm detection [S-A]
Algorithm detection	Earliest symptom to physician diagnosis [S-D]
Physician diagnosis (study index)	Algorithm detection to physician diagnosis [A-D]

HIPAA, Health Insurance Portability and Accountability Act; ICD-10, ICD-10 (International Classification of Diseases, 10th Revision); PH, pulmonary hypertension; WHO, World Health Organization; NLP, Natural Lanaguage Processing.

Cohort definitions

- PAH cohort: adult patients meeting PAH criteria (**Figure 1**)

Exposure

- Early detection of PAH using an ML algorithm, defined as the application of a previously developed ML model that analyzes structured and unstructured data to identify patients with a high probability of PAH based on combinations of clinical features, prior to actual physician diagnosis.

Outcomes

- DMCs saved in total and per patient per month (PPPM) for each resource type, which include visits, imaging, procedures, labs, and medications
- Measured 12 months pre-index and defined as the cost incurred during "*algorithm detection to physician diagnosis*" period, which is the potential cost saved due to the earlier detection by the algorithm

Statistical analyses

- A previously developed CatBoost ML model[2] was then employed to detect PAH earlier in the patient journey within 1 year of the earliest PAH symptom
- Unit costs derived from the Medicare Physician Fee Schedule[3] were multiplied with event counts to calculate total cost
- The per month cost for each patient was calculated as:

$$\text{Per month DMC saved} = \frac{\text{Cost saved due to earlier detection (incurred during A-D) (in USD)}}{\text{Time saved due to earlier detection (A−D) (in months)}}$$

- A subgroup analysis was also performed to compare patients with savings in the top and bottom quartiles

Table 1. Care setting–level event counts and associated total and PPPM costs incurred and saved for the PAH cohort

PAH cohort event counts and costs incurred in USD, N = 592			
PAH cohort total event counts			
Resource type	**Events from earliest symptom to physician diagnosis**	**Events from earliest symptom to algorithm detection**	**Estimated number of events averted**
Visits (office, remote, ED, IP, critical care)	4,532	874	3,658
Imaging (CT, MRI, x-ray, ECG, echo, VQ scan)	3,298	581	2,717
Procedures	1,067	139	928
Labs	16,994	3,901	13,093
Medications (oxygen supplementation)	336	57	279

PAH cohort total costs saved, USD			
Resource type	**Cost from earliest symptom to physician diagnosis**	**Cost from earliest symptom to algorithm detection**	**Estimated total costs saved**
Visits (office, remote, ED, IP, critical care)	5,541,208	1,190,056	4,351,152
Imaging (CT, MRI, x-ray, ECG, echo, V/Q scan)	430,568	58,216	372,353
Procedures	455,205	51,928	403,276
Labs	344,775	71,470	273,305
Medications (oxygen supplementation)	1,045	177	868
Total cost	6,772,800	1,371,847	5,400,953

PAH cohort PPPM costs saved, USD			
Resource type	**PPPM cost from earliest symptom to physician diagnosis**	**PPPM cost from earliest symptom to algorithm detection**	**Estimated PPPM costs saved[a]**
Inpatient visits (ED, IP, critical care)	4,960	592	4,455
Outpatient visits (office, remote)	297	31	273
Imaging (CT, MRI, x-ray, ECG, echo, V/Q scan)	1,090	76	1,022
Procedures	1,217	30	1,192
Labs	788	79	714
Medications (oxygen supplementation)	2	0	2
Total cost PPPM	8,355	808	7,658

[a]PPPM cost from earliest symptom to physician diagnosis and PPPM cost from earliest symptom to algorithm detection reflect average costs over different observation periods and are not directly subtractable. PPPM saved cost reflects the cost saved PPPM.

CT, computed tomography; ECG, electrocardiography; echo, echocardiography; ED, emergency department; IP, inpatient; MRI, magnetic resonance imaging; (V/Q), Ventilation/Perfusion.

- The final cohort included 592 patients. The mean age of patients was 62.7 ± 14.4 years, and 59% were females
- Time from symptom onset to detection/diagnosis
  - The mean time from earliest symptom to algorithm detection was 0.6 ± 2.2 months, compared with 4.0 ± 4.3 months from earliest symptom to physician diagnosis
- Estimated economic impact of earlier detection
  - Earlier detection resulted in mean DMC savings of $7,658 PPPM
- Primary drivers of cost savings

The primary contributors to estimated DMC savings resulting from earlier detection were:

  - Inpatient, emergency department, and critical care visits: $4,455 PPPM
  - Diagnostic imaging procedures: $1,022 PPPM
  - Outpatient, office, and remote visits: $273 PPPM
- Subgroup analysis (top vs bottom quartile of savings)
  - Patients in the top quartile of DMC savings were detected by the algorithm 3.7 months earlier compared with those in the bottom quartile of savings
  - This translated to approximately $13,540 PPPM greater DMC savings among patients in the top quartile

Limitations

- Potential for data missingness and coding or data entry errors within the EHR
- Data were derived from a single center, which may limit transferability and generalizability of the findings
- The performance and generalizability of the ML algorithm may be limited when applied to other health care settings, patient populations, or EHR systems

Conclusions

- Earlier detection of PAH using an ML algorithm was associated with meaningful estimated savings in pre-diagnosis DMC, averaging $7,658 PPPM. While this average savings may appear modest, it reflects that this analysis captures only pre-diagnosis costs averted through earlier detection rather than the full downstream costs of PAH management.
- Cost savings were primarily driven by reduced utilization of high-cost health care services (inpatient, emergency department, and critical care)
- Implementation of ML-based detection tools in health systems may support more efficient resource allocation and contribute to cost savings for patients with PAH

References

1. Didden E-M, et al. *Pulm Circ.* 2023;13(1):e12188.
2. Thakur T, et al. *A machine learning framework for early recognition of pulmonary arterial hypertension (PAH) using electronic-health record data.* Poster presented at ATS 2026 International Conference; May 17, 2026; Orlando, FL.
3. Centers for Medicare & Medicaid Services. PFS national payment amount file. https://www.cms.gov/medicare/payment/fee-schedules/physician/national-payment-amount-file

Disclosures

This study is sponsored by Merck Sharp & Dohme LLC, a subsidiary of Merck & Co., Inc., Rahway, NJ, USA.

Presented at International Society for Pharmacoeconomics and Outcomes Research (ISPOR); May 17-20, 2026; Philadelphia, PA, USA.