Accuracy of ICD-10 Codes for Identifying Pulmonary Embolism in Electronic Records: Yale New Haven Health System Validation for the PE-EHR+ Study

Shiwani Mahajan, MD, MHS; Jose Victor Jimenez Ceja, MD; César Caraballo, MD; Antoine Bejjani, MD; Candrika Khairani, MD; David Jimenez, MD; Eric Secemsky, MD; Frederikus Klok, MD, PhD; Andetta Hunsaker, MD; Mohamad Hussain, MD, PhD; Rohan Khera, MD, MS; Samuel Goldhaber, MD; Li Zhou, MD, PhD; Liqin Wang, PhD; Hamid Mojibian, MD; Gregory Piazza, MD; Harlan Krumholz, MD, SM; Behnood Bikdeli, MD, MS

Yale Traditional Internal Medicine Residency Program, New Haven, Connecticut

♦ Background

- The validity of the International Classification of Disease, 10th modification (ICD-10) codes used for identifying patients with pulmonary embolism (PE) is uncertain.
- The PE-EHR+ study[1] has been designed to validate the Principal Discharge Diagnosis (PDD) and the Secondary Discharge Diagnosis (SDD) ICD-10 codes used for PE.
- The study was conducted using data from Mass General-Brigham Health System, but validating the findings using data from another large health system is important.

♦ Objective

- To validate the results of the PE-EHR+ study using data from the Yale New Haven Health System (YNHHS).

♦ Methods

- 3 groups of patients in equal distribution (N=85 each) were identified from the YNHHS from 1/2016 to 12/2021, including those:
  - With ICD-10 PDD codes for PE
  - With ICD-10 SDD codes for PE
  - Without ICD-10 codes for PE
- The reference standard for identification of PE was manual chart review by two independent physicians using pre-specified criteria[1].
- Sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) were calculated.
- Weighted estimates were obtained by considering the total of number of hospitalizations at YNHHS in each group during the study period:
  - 3,697 patients with PDD codes for PE
  - 19,988 patients with SDD codes for PE
  - 294,326 patients without codes for PE

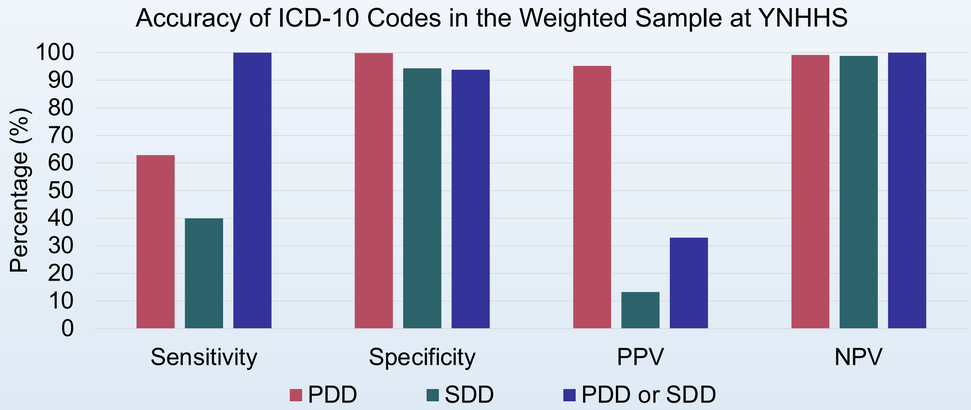
♦ Results

- 254 of the 255 manual charts reviewed were included (mean age: 63.3±18.9 years, 56.7% female) in the final analyses.
- Of these, 80 patients had a PDD code, and 10 patients had SDD code for PE.

Table. Accuracy of ICD-10 Discharge Diagnosis Codes for Identification of PE from YNHHS in the Unweighted Sample

	PDD	SDD	PDD or SDD
Overall, n	84	85	169
True positive, n	80	10	90
False negative, n	10	80	0
True negative, n	160	89	85
False positive, n	4	75	79
Total population	254	254	254
Sensitivity, %	80/90 = 88.9%	10/90 = 11.1%	90/90 = 100%
Specificity, %	160/164 = 97.6%	89/164 = 54.2%	85/164 = 51.8%

Note: True positive defined as patients with acute PE who were captured by Primary or Secondary Discharge Diagnosis for PE; False negative defined as patients with acute PE who were not captured by Primary or Secondary Discharge Diagnosis for PE; True negative defined as patients without acute PE who were not captured by Principal or Secondary Discharge Diagnosis codes for PE; False positive defined as patients without acute PE who were captured by Principal or Secondary Discharge Diagnosis codes for PE.



♦ Summary

- Using only the PDD codes for PE, the sensitivity and specificity of ICD-10 codes was 88.9% and 97.6% in the study sample, whereas the sensitivity and PPV of ICD-10 codes in the weighted analysis were 62.9% and 95.2% respectively.
- Incorporation of SDD codes to the PDD codes improved the weighted sensitivity to 100% but decreased the PPV to 33%.

♦ Conclusions

- The PDD codes miss a third of new PE events in hospitalized patients, though they have an excellent PPV.
- Integration of SDD codes improved sensitivity at the cost of worsening PPV.
- These results are consistent with findings from the Mass General-Brigham Health system (weighted sensitivity and PPV were 58.1% and 91.9% when using PDD, and 83.3% and 79.2% when using PDD or SDD).
- This highlights the need for modified strategies, such as use of ICD-10 discharges diagnosis codes in combination with procedure codes, or use of natural language processing–when feasible–for identification of patients with PE.

♦ References

1. Bikdeli B, Lo YC, Khairani CD, et al. Developing Validated Tools to Identify Pulmonary Embolism in Electronic Databases: Rationale and Design of the PE-EHR+ Study. *Thromb Haemost*. 2023;123(6):649-662.

♦ Questions?

- Please email shiwani.mahajan@yale.edu or bbikdeli@bwh.harvard.edu.