Developing and Evaluation of Computational Phenotypes of Metastatic Breast Cancer Using All of Us Data

Chenyu Li, MS, Israel O. Dilan, Shyam Visweswaran, MD PhD,
Michael J. Becich, MD PhD, Xia Jiang, PhD , Richard D. Boyce, PhD

Department of Biomedical Informatics, University of Pittsburgh School of Medicine

University of Pittsburgh

INTRODUCTION

Breast cancer is the most frequent cancer among women worldwide,[1] the metastatic breast cancer is the main cause of mortality for breast cancer patients. Computational phenotyping, an informatics approach, that extracts phenotypes from real-world data such as electronic health records (EHRs), has the potential to advance medicine's understanding of who is most at risk for metastatic breast cancer.
All of US (AoU) is a research program funded by the National Institutes of Health that aims to recruit one million highly diverse patients across the United States.[2] The program makes data from participants accessible to both participants and approved researchers through the AoU "workbench."

OBJECTICS

- Use AoU research workbench to implement the metastatic breast cancer phenotypes from multiple sources
- Describe cases within this rapidly growing a real clinical research database

METHODS

We applied two different computational phenotypes in December 2021:
- *Phenotype 1* was published on the PheKB knowledgebase of computational phenotypes by eMERGE group[3].
- *Phenotype 2* was from a classification and regression tree (CART) metastatic breast cancer phenotyping algorithm developed from a journal paper[4].

Phenotype 1 Definition from PheKB
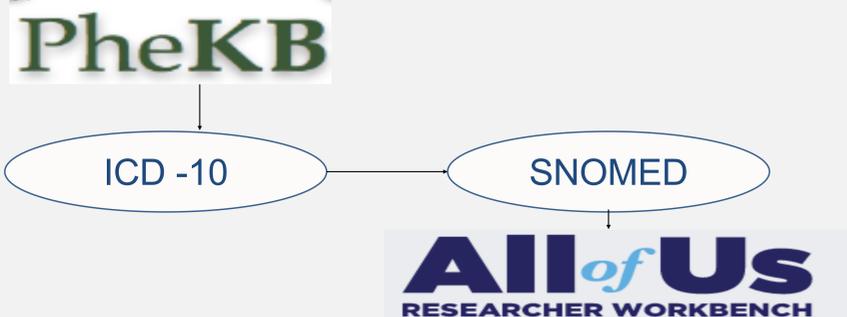
Phenotype 2 A CART model for metastatic breast cancer.
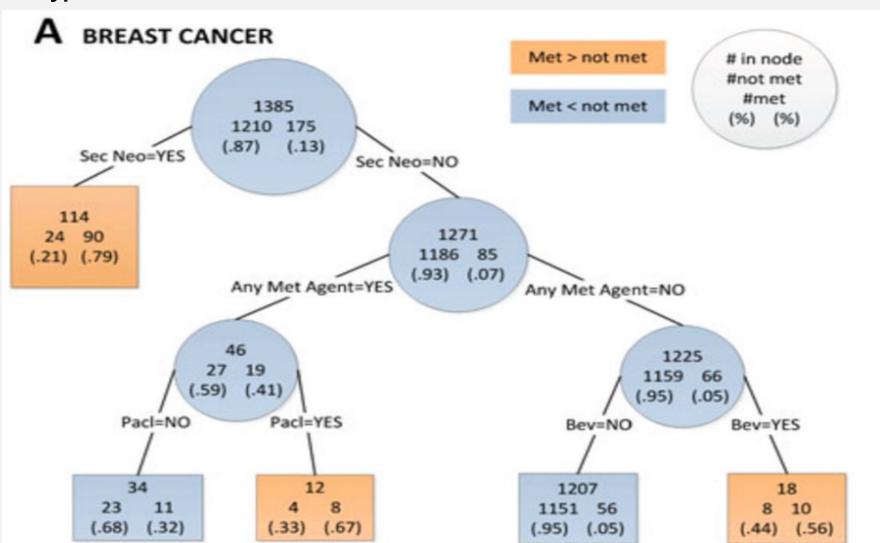
1. Secondary neoplasm = yes
2. Secondary neoplasm = no *and* any metastatic agent = yes *and* paclitaxel = yes
3. Secondary neoplasm = no *and* any metastatic agent = no *and* bevacizumab = yes

Figure 1 Flowchart for extracting two phenotypes

ACKNOWLEDGEMENTS

REFERENCES

1. Ferlay J et al. Global Cancer Observatory: Cancer Today. Lyon: International Agency for Research on Cancer; 2020
2. Denny JC, Rutter JL, Goldstein DB, et al. The "All of Us" Research Program. N Engl J Med 2019;381(7):668-
3. Shang Nea. Breast Cancer PheKB. Secondary Breast Cancer PheKB 2018. https://phekb.org/phenotype/1052.
4. Nordstrom BL, Whyte JL, Stolar M, Mercaldi C, Kallich JD. Identification of metastatic cancer in claims data. Pharmacoepidemiology and Drug Safety 2012;21:21-28 doi: 10.1002/pds.3247[published Online First: Epub Date]

CONTACT   chenyu.li@pitt.edu   https://www.linkedin.com/in/chenyu-li-80375196/

RESULTS

Of 201,920 participants in the AoU EHR database,
- Phenotype 1 identified 6,957 breast cancer patients while Phenotype 2 identified 3,679 patients.
- A small subset of these cohorts had metastatic breast cancer, 20 (0.28%) for Phenotype 1 and 190 (5.4%) for Phenotype 2.
- Only 8 participants were identified by both phenotypes.
- Phenotype 1, which used AoU workbench cohort construction tool SNOMED search, identified residents from only 3 states. While Phenotype 2, which queried ICD and CPT codes on workbench Python environment, identified residents from 7 states
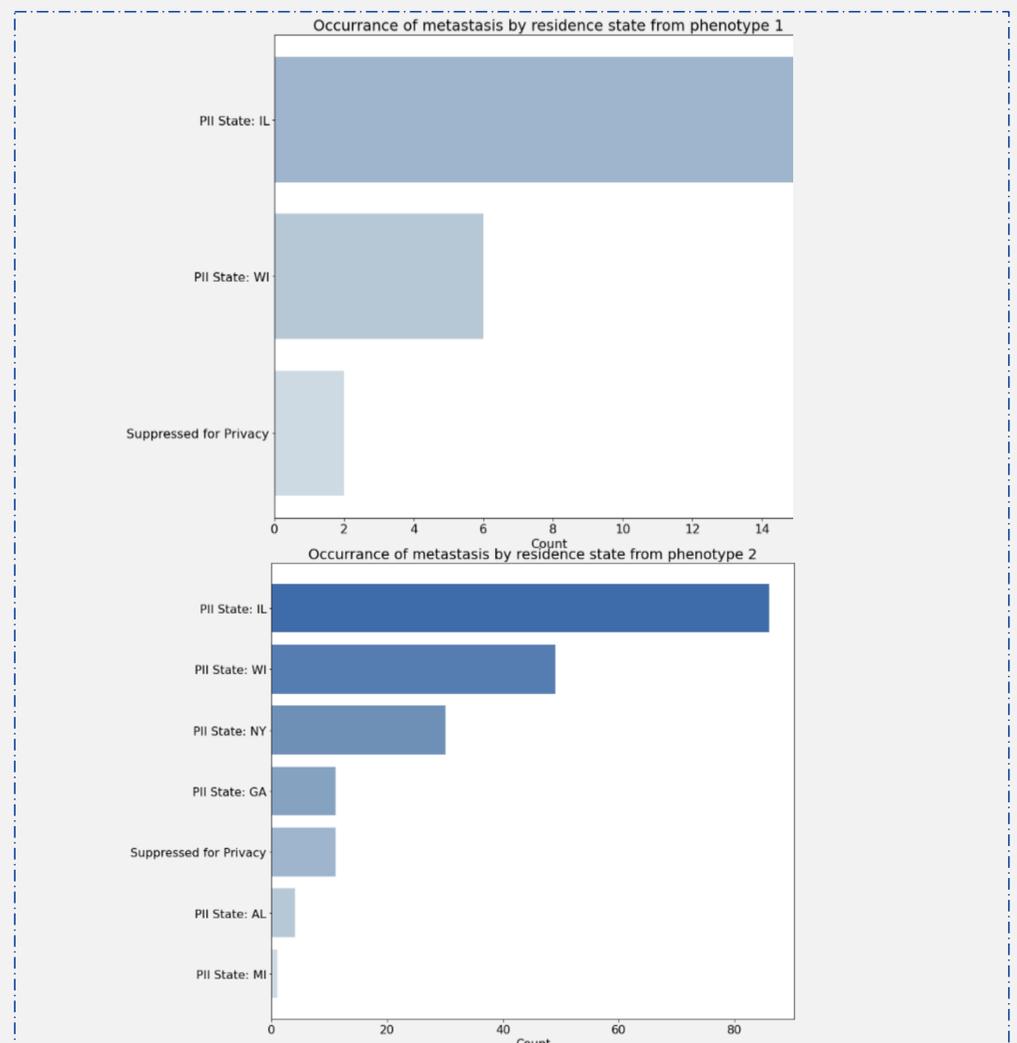

Figure 2 Occurrence of Metastatic Breast Cancer Patient Counts by Residence States

Breast cancer	Phenotype 1	Phenotype 2
Non metastatic	6,937	3,489
Metastatic	20	190
Total	6,957	3,679

Table 1 Number of non-metastatic and metastatic breast cancer AoU participants identified by each phenotype

More results

DISCUSSION

This study is the first to apply metastatic breast cancer phenotypes on AoU data.
- There were considerable differences in the number of metastatic patients identified by the two computational phenotype algorithms.
- AoU data contribution centers harmonized the EHRs data to the OMOP CDM, the data mapping process may vary from center to center.
- Researchers should be aware that concept coding will influence the participants capture.
- We suggest that comprehensive definition and detailed phenotyping algorithms of computational phenotypes should be reported in AoU data research.

FUTURE PLAN

In the future, we plan to assess how the AoU breast cancer population differs from other established databases and how different phenotypes will influence metastatic breast cancer prediction models.