INTRODUCTION

Breast cancer is the most frequent cancer among women worldwide, 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. The program makes data from participants accessible to both participants and approved researchers through the AoU workbench.

OBJECTIVES

• Use AoU research workbench to implement the metastatic breast cancer phenotypes from multiple sources
• Describe cases within this rapidly growing 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.
• Phenotype 2 was from a classification and regression tree (CART) metastatic breast cancer phenotyping algorithm developed from a journal paper.

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.

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.

ACKNOWLEDGEMENTS

This study was partially supported by the U.S. Department of Defense Award No. W81XWH1910495 (to XJ) and NIH OT2 OD026554. Other than supplying funds, the funding agencies played no role in the research.

REFERENCES


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.