

Plan Overview

A Data Management Plan created using DMPTool

DMP ID: <https://doi.org/10.48321/D1107X>

Title: Virtual Clinical Trial of Contrast-Enhanced Tomosynthesis for Screening of Dense Breasts

Creator: Paul Kinahan - **ORCID:** [0000-0001-6461-3306](https://orcid.org/0000-0001-6461-3306)

Affiliation: University of Washington (uw.edu) (washington.edu)

Principal Investigator: Cynthia Davis, Paul E Kinahan

Data Manager: Cynthia Davis, Paul E Kinahan

Project Administrator: Paul E Kinahan

Funder: National Institutes of Health (nih.gov)

Funding opportunity number: PAR-21-166

Template: NIH-Default DMSP

Project abstract:

Contrast-enhanced (CE) imaging has demonstrated improvements for breast cancer detection in women with dense breasts. While CE-MRI is potentially a gold standard, it is not accessible as a widespread screening approach, so a priority for screening is the evaluation of contrast enhancement with 2D mammography (CE-M) and 3D digital breast tomosynthesis (CE-DBT). To compare the relative effectiveness of CE-M and CE-DBT in a timely and feasible manner, we are proposing an in silico 'virtual imaging clinical trial' that uses population-based and realistic models of breast anatomy, cancer biology, physics models of the image acquisition processes, and models of human observers.

Start date: 07-01-2024

End date: 06-30-2029

Last modified: 01-18-2024

Copyright information:

The above plan creator(s) have agreed that others may use as much of the text of this plan as they would like in their own plans, and customize it as necessary. You do not need to credit the creator(s) as the source of the language used, but using any of the plan's text does not imply that the creator(s) endorse, or have any relationship to, your project or proposal

Virtual Clinical Trial of Contrast-Enhanced Tomosynthesis for Screening of Dense Breasts

The primary type of scientific data generated (that will not be captured in publications) are synthetic medical images. The images will be stored the industry-standard Digital Imaging and Communications in Medicine (DICOM) standard. The synthetic images will be representations of contrast-enhanced mammography and contrast-enhanced digital breast tomosynthesis images.

Based on the VICTRE trial [Badano A, et al. JAMA Network Open. 2018], we expect to generate on the order of 30,000 contrast-enhanced mammography and contrast-enhanced digital breast tomosynthesis images (each) from approximately 3,000 virtual patients. E.g. 100 images of both types for each virtual patient.

The DICOM images will be approximately 50MB in size for each contrast-enhanced mammogram and 1 GB for a digital breast tomosynthesis image. The DICOM images will be stored on the Zenodo server at CERN.

Zenodo is a general-purpose open repository developed under the European OpenAIRE program and operated by CERN. For each data set, a persistent digital object identifier (DOI) will be created, which makes the stored items easily citeable and searchable. Zenodo addresses the FAIR Principles definition as referenced from: Wilkinson, M. D. et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* 3:160018 doi: 10.1038/sdata.2016.18 (2016).

The primary type of scientific data generated are synthetic mammography and digital breast tomosynthesis images. The images will be stored the industry-standard Digital Imaging and Communications in Medicine (DICOM) standard. This unique data, with ground truth metadata, will be made available for other researchers for developing and/or testing image processing algorithms to be used with mammograms and digital breast tomosynthesis images.

To facilitate the interpretation and reuse of the data, a README file and data dictionary will be generated and deposited into a repository along with all shared datasets. The README file will include synthetic patient and lesion information, scanner descriptions, and acquisition protocol. The data dictionary will define and describe all variables in the dataset.

The image data generated via the simulations will be in the Digital Imaging and Communications in Medicine (DICOM) format, which is well-defined and widely adopted. Thus free and commercial tools to read and view the images are readily available.

Question not answered.

Imaging data and metadata will be deposited into Zenodo.

We will use Persistent Unique Identifiers (PIDs) to improve data findability across all dissemination outputs. PIDs used will DOIs for outputs (e.g., datasets, protocols), Research Resource IDentifiers (RRIDs) for resources, and Research Organization Registry (ROR) IDs and funder IDs for places, as much as possible to make data identifiable and findable. We will also use indexed metadata, such as MeSH terms with a unique URL to make scientific data easily findable. We will keep our records up to date with DOIs for our datasets and publications, ROR, and funder IDs to increase findability.

All scientific data generated from this project will be made available as soon as possible, and no later than the time of publication or the end of the funding period, whichever comes first. The duration of preservation and sharing of the data will be a minimum of 10 years after the funding period.

There are no factors that will impact access, distribution, and reuse for all other scientific data generated by this study.

Access to scientific data will not be controlled.

Question not answered.

The schema for metadata and storage of the images data will be based on the core Digital Imaging and Communications in Medicine (DICOM) standard, which will be the format the images are stored in. The schema will be proposed and reviewed by the leadership group. Once the schema is finalized, sample data will be uploaded to Zendo to test adherence to the FAIR principles, and the schema will be modified if necessary.

As the research data sets are created, they will undergo a two step process between the University of Washington and GE-Research teams. A member from one team will upload the data set, and a member of the other team will verify that that data are findable, downloadable, and correct.
