Plan Overview

A Data Management Plan created using DMP Tool

DMP ID: https://doi.org/10.48321/D1MT0T

Title: ProsGATE: Geriatric Assessment and Technology Evaluation in Prostate Cancer

Creator: Nabiel Mir - **ORCID:** 0000-0003-0635-3985

Affiliation: University of Chicago Medical Center (uchospitals.edu)

Project Administrator: Ryan Priest, Michelle Jarvis , Melissa Vonesh

Funder: National Institutes of Health (nih.gov)

Funding opportunity number: PAR-22-227

Grant: https://grants.nih.gov/grants/guide/pa-files/PAR-22-227.html

Template: NIH-Default DMSP

Project abstract:

Management of metastatic prostate cancer (mPCa) in older adults is challenging, with frailty influencing treatment tolerance and outcomes. Despite guidelines recommending a Comprehensive Geriatric Assessment (GA) for frailty evaluation, less than half of older adults with cancer undergo this evaluation in practice due to the cumbersome nature of current tools. The Geriatric Remote Initiative (GeRI) aims to streamline frailty assessments by integrating digital health technologies—such as accelerometers, bioimpedance devices, and patient-reported outcomes—into cloud-based risk models. Early pilot data has confirmed high user acceptance and usability of the GeRI platform in older adults with mPCa. This project seeks to expand GeRI's capabilities to predict treatment-related morbidity, offering personalized and data-driven care for older adults with mPCa. This proposal focuses on assessing wearable device-derived metrics to predict frailty and adverse outcomes in older adults with mPCa receiving androgen deprivation therapy (ADT). In collaboration with industry partner EMA Wellness, we will evaluate the predictive value of step count and body composition in relation to clinical frailty and symptom burden, aiming to improve treatment risk stratification in mPCa. The project is built on three specific aims: (1) Validate the correlation between Samsung Galaxy Watch (SGW) step counts and Actigraph measurements, hypothesizing a strong correlation ($\rho \geq 0.7$) over seven days of monitoring; (2) Determine if step counts, lean, and fat mass correlate with frailty and symptom burden (pain and fatigue) at baseline and during ADT treatment, hypothesizing moderate correlations ($\rho \geq 0.4$); and (3) Evaluate the predictive power of baseline digital biomarkers in forecasting serious adverse events (SAEs) during treatment, positing that step count, lean mass, and fat mass

will outperform traditional frailty measures in predicting SAE incidence. Through these aims, this project aims to establish validated digital biomarkers that can facilitate remote frailty assessment and personalized treatment management in older adults with mPCa. The REDI K01 award will support the development of predictive models using digital health data, furthering my transition to an independent physician-scientist focused on integrating digital biomarkers into oncology care.**

Start date: 07-02-2024

End date: 07-02-2027

Last modified: 10-17-2024

Copyright information:

ProsGATE: Geriatric Assessment and Technology Evaluation in Prostate Cancer

Data Type

Types and amount of scientific data expected to be generated in the project: *Summarize the types and estimated amount of scientific data expected to be generated in the project.*

Describe data in general terms that address the type and amount/size of scientific data expected to be collected and used in the project (e.g., 256-channel EEG data and fMRI images from ~50 research participants). Descriptions may indicate the data modality (e.g., imaging, genomic, mobile, survey), level of aggregation (e.g., individual, aggregated, summarized), and/or the degree of data processing that has occurred (i.e., how raw or processed the data will be)

This project will generate biometric, body composition, clinical frailty, and patient-reported outcome data from wearable devices (Samsung Galaxy Watch), bioimpedance assessments, and self-reported surveys (e.g., PRO-CTCAE measures). Data will be collected from 60 older adults (≥65 years) with metastatic prostate cancer (mPCa) undergoing androgen deprivation therapy (ADT). Additionally, a validation cohort of 20 older adults will be used in Aim 1 to assess the agreement between step counts captured by the Samsung Galaxy Watch (SGW) and Actigraph accelerometers. Data from this validation study will include side-by-side step count measures over 7 days of continuous monitoring. Three primary datasets will be generated, estimated at approximately TBD. These datasets will include: (1) raw step count and bioimpedance data (lean and fat mass) collected via Samsung Galaxy Watch for both the validation and primary study cohorts, (2) clinical frailty data from an ASCO-compliant 6-point Geriatric Assessment scale, and (3) 4-point patient-reported outcomes on pain and fatigue. Serious Adverse Events (SAEs) will be recorded and monitored by the investigative team. Intermediate files will include cleaned and merged datasets consolidating biometric data with clinical frailty assessments and patient-reported outcomes. Post-processed files will contain data ready for statistical analysis, segmented by study aims. Raw data will undergo transformations using statistical methods such as mixed-effects modeling, ordinal regression, and machine learning algorithms (e.g., Random Forests, XGBoost). These analyses will explore correlations and predictive relationships between digital biomarkers and treatment outcomes. All datasets will be de-identified and aggregated for sharing, protecting participant confidentiality. Data collection will be conducted at clinical sites in the Chicago area, with remote biometric and patient-reported outcome data captured via wearable devices from participants. The validation cohort data will be crucial for determining the accuracy of the SGW in relation to research-grade accelerometers, ensuring reliable data for subsequent analysis in the broader cohort.

Scientific data that will be preserved and shared, and the rationale for doing so: *Describe which scientific data from the project will be preserved and shared and provide the rationale for this decision.*

Based on ethical and legal considerations, the following scientific data produced in the course of the project will be preserved and shared: aggregated and anonymized biometric data (step count, lean mass, fat mass), clinical frailty assessment scores, patient-reported outcomes (pain, fatigue), and summary statistics of Serious Adverse Events (SAEs). Individual participant identifiers and any other personally identifiable information will not be shared to protect participant privacy and comply with regulatory standards. The final dataset will include aggregated biometric data from wearable devices (Samsung Galaxy Watch), clinical frailty scores from the ASCO-compliant frailty scale, patient-reported outcomes on pain and fatigue, and clinical information related to treatment, including SAEs. Data de-identification practices will include the removal of all personally identifiable information, the use of unique coded identifiers, and data aggregation before sharing. These practices will be reflected in informed consent forms to ensure transparency with participants. The rationale for sharing these specific subsets of data is to contribute to the broader scientific understanding of the relationships between digital biomarkers, frailty, symptom burden, and treatment outcomes in older adults with metastatic prostate cancer (mPCa) undergoing androgen deprivation therapy (ADT). Sharing de-identified, aggregated data will allow for external validation of findings, foster collaborative research, and support the wider dissemination of knowledge.

Ultimately, this could contribute to improved clinical care, treatment personalization, and outcomes for older adults with mPCa, while ensuring compliance with ethical and legal standards governing human subjects research.

Metadata, other relevant data, and associated documentation: Briefly list the metadata, other relevant data, and any associated documentation (e.g., study protocols and data collection instruments) that will be made accessible to facilitate interpretation of the scientific data.

To facilitate the interpretation of the data, a comprehensive data dictionary, metadata, documentation, statistical analysis plans, study protocols, and data collection instruments will be created, shared, and associated with the relevant datasets. In addition to the aggregated and anonymized data being shared, the final data package will include the statistical analysis code, analytic plans, transformation algorithms used during data processing, and any machine learning models employed (e.g., XGBoost, Random Forest). Documentation and support materials will be compatible with the clinicaltrials.gov Protocol Registration Data Elements. The data dictionary will provide detailed definitions and explanations for each variable in the datasets, including methods used for data collection, such as step count and bioimpedance data from the Samsung Galaxy Watch, frailty measures, and patient-reported outcomes (pain, fatigue). Metadata will describe the data collection process, cleaning and transformation procedures, and the statistical methodologies used for analysis. Study protocols will detail the design, inclusion/exclusion criteria, methodology, and ethical considerations. Data collection instruments, including assessment forms, surveys, and wearable device interfaces, will be shared to clarify how the data was collected. A README file will accompany the datasets, offering a structured overview of the data package, metadata, and documentation, with guidance on how to interpret and use the data accurately.

Related Tools, Software and/or Code

State whether specialized tools, software, and/or code are needed to access or manipulate shared scientific data, and if so, provide the name(s) of the needed tool(s) and software and specify how they can be accessed.

Patient-Reported Outcome (PRO) and biometric data will be made available in a proprietary format that requires specialized tools for access and manipulation, such as the EMA Wellness app and associated software. These tools are proprietary, fee-based software provided by the EMA Wellness team and are essential for managing and analyzing the collected data. The software ensures data integrity, accuracy, and provides a HIPAA-compliant data flow and storage system vetted by institutional IT security. EMA Wellness has committed to supporting the longevity of these tools, including updates and secure deletion of data upon study completion. For broader accessibility, data will also be exported to a nonproprietary format (e.g., CSV), allowing for data use and reuse by investigators and the wider research community. This alternative format can be manipulated with widely available software tools such as Excel or R. Links to file viewers or open-source software options for accessing and manipulating the CSV data will be provided, along with comprehensive documentation to facilitate understanding and effective use of the data in this format. This dual approach ensures that researchers can utilize both the specialized proprietary tools for full-featured data analysis and the more accessible CSV format for broader data sharing and reuse.

Standards

State what common data standards will be applied to the scientific data and associated metadata to enable interoperability of datasets and resources, and provide the name(s) of the data standards that will be applied and describe how these data standards will be applied to the scientific data generated by the research proposed in this project. If applicable, indicate that no consensus standards exist

Data will be stored in common and open formats, such as CSV, for both Patient-Reported Outcome (PRO) data and biometric data. To enable interoperability, associated metadata will include information such as the meaning of variable names, coding schemes, handling of missing data, and references to the sources of standardized terms where applicable. A comprehensive data dictionary will accompany the datasets, ensuring consistent interpretation of variables and facilitating reuse by other researchers. Although formal standards for biometric data have not yet been widely adopted, the project will adhere to best practices, including the use of structured metadata, rigorous validation, and quality control processes. These practices will ensure data consistency and integrity. For PRO data, we will follow the Patient-Reported Outcomes Measurement Information System (PROMIS) guidelines where applicable, ensuring that any standardized names and terms align with existing frameworks for health-related data. Where consensus standards do not exist, the data will be documented and structured following best practices to ensure ease of integration with future standards and compatibility with a wide range of analysis platforms.

Data Preservation, Access, and Associated Timelines

Repository where scientific data and metadata will be archived: Provide the name of the repository(ies) where scientific data and metadata arising from the project will be archived.

Specialized data, such as Electronic Patient-Reported Outcomes (ePROs), step count, and body composition data collected via the EMA Wellness app and Samsung Galaxy Watch, will be securely stored in EMA Wellness's proprietary repository. This repository is designed for HIPAA-compliant data storage and sharing, validated by our IT security department. De-identified data will be shared with investigators and extracted for further analysis using REDCap Case Report Forms, which will be securely stored on the UChicago Center for Research Informatics (CRI) servers.

Clinical data, including demographic information, cancer/treatment-related data, geriatric assessments, and treatment outcomes (e.g., serious adverse events), will be archived using REDCap on the CRI cloud. This platform ensures robust data preservation, long-term access, and discoverability, making it ideal for storing and sharing scientific data securely.

How scientific data will be findable and identifiable: Describe how the scientific data will be findable and identifiable, i.e., via a persistent unique identifier or other standard indexing tools.

Clinical data, including demographic, clinical/cancer/treatment, geriatric assessments, and treatment outcomes (e.g., serious adverse events), will be stored in the CRI cloud. The CRI cloud repository ensures that data is findable and identifiable via persistent unique identifiers such as Digital Object Identifiers (DOIs). Comprehensive metadata will accompany the datasets, making them accessible through standard indexing tools. The CRI repository is maintained by the University of Chicago, and datasets are available to UChicago Biological Science Division Faculty who meet the access requirements, with data requests being processed through a controlled dissemination process to ensure data security and compliance. The EMA Wellness proprietary repository will similarly provide metadata and persistent identifiers, such as DOIs, to ensure the findability and long-term accessibility of the data. Datasets stored in EMA Wellness's repository are accessible under a Master Services Agreement with the University of Chicago Medical Center, ensuring that access and sharing are HIPAA-compliant and governed by institutional agreements.

When and how long the scientific data will be made available: Describe when the scientific data will be made available to other users (i.e., no later than time of an associated publication or end of the performance period, whichever comes first) and for how long data will be available.

Shared data generated from this project will be made available upon request by interested parties no the funding period's end. The duration of preservation and sharing of the data will be a minimum of 10 years after the end of

the funding period.

Access, Distribution, or Reuse Considerations

Factors affecting subsequent access, distribution, or reuse of scientific data: NIH expects that in drafting Plans, researchers maximize the appropriate sharing of scientific data. Describe and justify any applicable factors or data use limitations affecting subsequent access, distribution, or reuse of scientific data related to informed consent, privacy and confidentiality protections, and any other considerations that may limit the extent of data sharing.

Due to ethical, legal, and privacy considerations, access, distribution, and reuse of the resulting scientific data will be overseen by a designated committee. This committee will review all data-sharing requests to ensure compliance with informed consent agreements, privacy and confidentiality protections, and institutional policies. Requests for data access must be formally submitted, and the committee will evaluate each request on a case-by-case basis to ensure alignment with the University of Chicago's data sharing policies and NIH guidelines. The committee will also monitor the distribution and reuse of the data to maintain the integrity of the research and safeguard participant privacy. Data that includes sensitive or personally identifiable information will be carefully controlled to prevent any unauthorized access or misuse. In cases where full sharing may pose risks to participant confidentiality, only aggregated or de-identified datasets will be shared. This structured approach allows for appropriate data sharing while adhering to ethical, legal, and institutional frameworks.

Whether access to scientific data will be controlled: State whether access to the scientific data will be controlled (i.e., made available by a data repository only after approval).

Given the sensitive nature of the dataset, data will be made available in the CRI Cloud data repository, which restricts access to the data to qualified investigators with an appropriate research question and an approved Data Use Agreement (DUA). Data can be accessed through a formal request process, which entails submission of a detailed research proposal and the requisite Data Use Agreement to the designated review committee. Upon approval, investigators will be granted access to the data via a secure online platform, ensuring both the integrity of the data and compliance with established ethical and legal frameworks.

Protections for privacy, rights, and confidentiality of human research participants: If generating scientific data derived from humans, describe how the privacy, rights, and confidentiality of human research participants will be protected (e.g., through de-identification, Certificates of Confidentiality, and other protective measures).

In order to ensure participant consent for data sharing, IRB paperwork and informed consent documents will include language describing plans for data management and sharing of data, describing the motivation for sharing, and explaining that personal identifying information will be removed.

To protect participant privacy and confidentiality, shared data will be de-identified using the methods in compliance with HIPAA guidelines. This will include the removal or modification of any information that could be used to identify individuals, such as names, geographical information, and unique identifiers.

Moreover, access to the data will be governed by an approval process, where only qualified investigators with a valid research question and an approved Data Use Agreement (DUA) will be granted access. This controlled access mechanism, alongside the robust de-identification process, fortifies the safeguarding of participant privacy, rights, and confidentiality, fostering a conducive and ethically sound environment for the conduct and dissemination of human-centric research.

Oversight of Data Management and Sharing

Describe how compliance with this Plan will be monitored and managed, frequency of oversight, and by whom at your institution (e.g., titles, roles).

Lead PI Nabiel Mir MD, ORCID: 0000-0003-0635-3985, will be responsible for the day-to-day oversight of lab/team data management activities and data sharing. Broader issues of DMS Plan compliance oversight and reporting will be handled by the PI and Co-I team as part of general campus(es) stewardship, reporting, and compliance processes. The compliance with the Plan will be reviewed on regularly by a designated Data Safety Monitoring Committee within the institution. The committee will evaluate adherence to the data management and sharing procedures outlined in the DMS Plan, assess the effectiveness of de-identification methods, and ensure that data sharing activities align with informed consent agreements. Any non-compliance issues identified will be addressed promptly, and corrective measures will be implemented as needed to ensure ongoing compliance with the DMS plan.

Planned Research Outputs

Dataset - "Biometric and Clinical Dataset of Older Adults with Metastatic Prostate Cancer "

This dataset stems from a multicenter, prospective cohort study involving 60 older adults with metastatic hormone-sensitive prostate cancer (mHSPC), spanning 12 months. It captures the patients' health trajectory under androgen deprivation therapy (ADT) plus androgen receptor signaling inhibitors (ARSI). Biometric data, including daily step count, lean mass, and fat mass, are measured using Samsung Galaxy Watch 5 and EMA Wellness software, providing a comprehensive longitudinal view of physical activity and body composition. Frailty assessments, conducted at baseline and at 3, 6, 9, and 12-month intervals, utilize Geriatric-8 (G8), Katz Activities of Daily Living (ADL), 4-item Instrumental ADL (iADL), Cumulative Illness Rating Scale-Geriatric (CIRS-G), Montreal Cognitive Assessment (MoCA), and Mini Geriatric Depression Scale (Mini-GDS).

The dataset records treatment regimens, dosages, frequencies, and the occurrence and classification of Serious Adverse Events (SAEs). Patient-reported outcomes (PROs) of pain and fatigue, graded per the PRO-CTCAE measures, are meticulously tracked. Demographic and cancer-specific data are included to provide rich context for analysis. The dataset is temporally structured to support nuanced longitudinal analysis.

It serves as a crucial resource for investigating the relationships between biometric indicators and clinical frailty, evaluating treatment impacts, and assessing the predictive power of digital biomarkers for adverse events in geriatric oncology. Data accessibility is managed through secure, HIPAA-compliant repositories with controlled access, ensuring privacy and facilitating collaborative research. Accompanying metadata and persistent identifiers enhance discoverability and access. Preservation and sharing protocols adhere to the funding institution's guidelines and ethical standards, ensuring the dataset’s long-term utility for the broader scientific community.

This dataset is poised to make significant contributions to geriatric oncology by exploring the dynamic interplay between physical activity, body composition, frailty, and treatment outcomes in older adults with mHSPC.

Planned research output details

Title	Type	Anticipated release date	Initial access level	Intended repository(ies)	Anticipated file size	License	Metadata standard(s)	May contain sensitive data?	May contain PII?
Biometric and Clinical Dataset of Older Adults wit ...	Dataset	2028-01-02	Restricted	None specified		None specified	None specified	No	No