Plan Overview

A Data Management Plan created using DMPonline

Title: Home based exercise for patients with breast or prostate cancer during treatment: a feasibility trial

Creator: Larissa Xavier Neves da Silva

Principal Investigator: Daniel Umpierre de Moraes, Daniel Umpierre de Moraes

Data Manager: Daniel Umpierre de Moraes, Daniel Umpierre de Moraes, Jayne Santos Leite, LARISSA XAVIER NEVES DA SILVA

Project Administrator: Daniel Umpierre de Moraes, Daniel Umpierre de Moraes, LARISSA XAVIER NEVES DA SILVA

Affiliation: Other

Template: DCC Template

ORCID iD: 0000-0001-6953-0163

ORCID iD: 0000-0001-6953-0163

Project abstract:

This a PhD project based on a feasibility trial to assess home-based exercise program, including a 3-month intervention of structured physical activity and health education for individuals with breast cancer or prostate. Our main objective is to assess the adherence of cancer patients to a program of structured exercise, mixed with a health education approach, implemented and accompanied in a remote approach. Therefore, our data collection will include variables related to: (a) recruitment and retentions rates; (b) attendance to exercise sessions; (c) fatigue and quality of life levels; (d) functional capacity levels; (e) adverse events related or not to the study; (f) patient selected outcomes.

ID: 73965

Start date: 22-03-2022

End date: 29-12-2023

Last modified: 31-03-2022

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 customise 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

Home based exercise for patients with breast or prostate cancer during treatment: a feasibility trial

Data Collection

What data will you collect or create?

1. Trial and intervention performance
1.1 Recruitment rates
1.2 Retention rates
1.3 Training compliance

2. Individual health-related data
2.1 Quality of life
2.2 Fatigue
2.3 Functional capacity
2.4 Anthropometry
2.5 Adverse events

3. Trial materials
3.1 Video and pictures of the training exercises (audiovisual data; PNG and JPEG files)
3.2 Educational content (text data; TXT and HTML files)
Considering all these data, we expect to generate a large volume of them, in Giga Bytes (GB), especially because of the video materials. No existing data will be reused. All the collected data will be uploaded to the Red Cap platform and also backed up in Google Documents, password protected. After the completion of the study and after the first article being published, analysis datasets and processed variables will be de-identified and shared in the Zenodo.org plataform. All relevant data will be saved and kept throughout the entire study and beyond (all the allowed time from the platform).

How will the data be collected or created?

Data collection
The data will be collected through a range of study instruments, which are listed below:

Recruitment
- Study questionnaire for eligibility screening (remote by phone contact)
- Written informed consents (applied either remotely or face-to-face; archived both in digital files and in paper-format whenever applicable)

Quality of life
- Questionnaire using a validated instrument (remote or face-to-face)

Fatigue
- Questionnaire using a validated instrument (remote or face-to-face)

Anthropometry

- Body measurements using a validated protocol (face-to-face)

Functional capacity

- Walking test using a validated protocol (face-to-face)

Adverse events

- Study questionnaire for self-reported adverse events (remote or face-to-face)

Intervention resources

- Video, pictures, and text related to intervention will be sources from openly licensed materials or created by the study team, under a Creative Commons License 4.0.

Standardised procedures

In all possible scenarios, we will use Case Report Forms (CRF) in an online manner (e.g., Google Forms Sheet), which will allow standardised data capture, as well as facilitate typing, versioning, or uploading of documents.

In addition, each assessment will have a standardised operational procedure (SOP) to increase internal consistency. Such SOPs will determine who may conduct the assessment (either online or face-to-face), evaluation steps, and standardised communication with research participants.

Data harmonization

Whenever possible, we will use reference terminology to code the study outcomes. This will be made through the NCI Thesaurus, which covers vocabulary for clinical care, translational and basic research, and public information and administrative activities.

Data entry and manipulation

To allow consistency analyses, we will use a duplicate approach to carry out the data entry, using working platforms such as RedCap. Although this procedure will be standard for all data entry, possible exceptions may apply. In such cases, the data will be annotated as being single-typed, which seeks to facilitate future data verification or quality control routines.

Data manipulation to create new variables, merging existing ones, or changing anything that may be needed will be made through command-line software such as RStudio or with written annotation of changes. In such cases, new files or datasets will be created so that the original ones may be preserved readily accessible.

Folders and files

Our folders will indicate whether the archived data refers to:

- regulatory documents (e.g., project versions, amendments, IRB communications)
- study materials (e.g., informed consents, assessment forms, intervention resources)
- generated research participants' data (e.g., CRFs, raw sheet, datasets after analyses)
- dissemination documents (e.g., individual result reports, trial results reports, trial reports/articles)

Files will be named in English in the format as follows: "filename yyyy-mm-dd" (e.g., "crf_walking_test_2021_04_20").

Documentation and Metadata

What documentation and metadata will accompany the data?

We will provide data documentation to shared files containing individual participant data. The annotations will include:

1. Date
2. Version (with release updates)
3. Creators
4. Data dictionary (standardised vocabulary will be used for existing terms at the NCI Thesaurus)
5. Persistent identifier (DOI)
6. Licensing
7. Citation (credits)

Ethics and Legal Compliance

How will you manage any ethical issues?

Data preservation and sharing

Research participants will be asked whether they authorize use to keep and share their individual, de-identified data for research use only. Those who do not allow us to share their de-identified data will be removed from the project dataset(s) to be publicly shared.

De-identification

Datasets will be handled without any information on subjects' names, town of birth, town of residence, date of birth, other personal dates (e.g., date of diagnosis), zip code, phone number, email address. Moreover, files will be as de-identified as possible before public data sharing. Although unlikely due to the prevalence of breast and prostate cancer, personal information that may allow identification includes age, sex, race/color, number of offspring (if any), conjugal status, type of diagnosis, duration of diagnosis, socioeconomic status, city of residence.

Ethical consultation

For issues not anticipated at the time of redacting the present plan, we will consult the Division of Bioethics and the Institutional Ethics Committee at the Hospital de Clínicas de Porto Alegre (Porto Alegre, Brazil).

How will you manage copyright and Intellectual Property Rights (IPR) issues?

Structured data that will result from this research project will pertain to the PI (Daniel Umpierre), who will be responsible for data integrity. After data protection and data handling routines, the de-identified data will be readily available to all project contributors.

Each folder or dataset to be publicly shared will display the licensing and copyright information.

Finally, we consider that research data gathered from data collection from research participants pertain primarily to these individuals. Therefore, all individual data will be available to their owners

throughout the project.

Storage and Backup

How will the data be stored and backed up during the research?

Data protection
This project will generate mostly sensitive data. Therefore, every month we will generate local (machine) copies of files containing online datasheets (e.g., data retrieved from Google forms) containing individual participant data. These files (mostly xlxs files) will be uploaded to: (1)a RedCap repository with daily automated backups and (2) to an institutional Google Drive folder ("generated research participants' data", mentioned above) with restricted access.

Data handling
The participants will have their identity recorded in their personal identification CRF, together with residential address and means of contact (phone number(s) and email). Their name will also appear in the written informed consent. Both documents will be kept in a separated institutional Google Drive folder accessible by any project contributor. The manipulation of these files (editor roles) will be possible to the PI, and strictly authorised personnel. At the time of this plan, two other people are allowed to edit documents: MsC. Larissa Xavier Neves da Silva and MsC. Jayne Santos Leite.

How will you manage access and security?

All the files will be password-protected and only investigators or data managers designated to use the files will be given access. Different passwords will be used for different users, which allows auditing to be made more easily. Although we do not plan to request justification to release the datasets, we do plan to share the data in a controlled approach. Therefore, once the datasets are made available at the Zenodo.org platform, only registered users will be able to download files. This procedure allows us to keep a record of data requests.

Selection and Preservation

Which data are of long-term value and should be retained, shared, and/or preserved?

Due to space and management constraints, we are planning:
1. To retain and store project documentation (informed consents, IRB communication, subjects' communication) for at least 10 years after the study completion (anticipated time: up to 2033).

2. To retain and share project resources (instruments, intervention materials) as long as public repositories (primarily OSF; and, alternatively, Zenodo) provide sharing services at no cost.

3. To retain and share datasets as long as public repositories (primarily Zenodo; and, alternatively, OSF) provide sharing services at no cost.

Note: expected preservation for Zenodo is of at least 20 years from the beginning of this resource (therefore, until 2033). Concerning the OSF, their funding is sufficient to maintain read access to hosted data for nearly 50 years.

Links regarding longevity plans for Zenodo and OSF:
1. https://about.zenodo.org/
2. https://help.osf.io/hc/en-us/articles/360019737894-FAQs

What is the long-term preservation plan for the dataset?

We plan to retain the data for the available time that is set at the Zenodo.org platform. Since our final files will be stored either at Zenodo (datasets) and OSF (materials and resources), we expect that datasets will be available for, at least, one decade (until 2033, hosted at Zenodo).

Data Sharing

How will you share the data?

The data will be shared using two public repositories: OSF (https://osf.io/) and Zenodo (https://zenodo.org/). The OSF will be used for project resources that **do not** contain individual participant data, whereas Zenodo will be used to shared datasets containing data obtained from research participants. The reason for this choice is that Zenodo allows a "Restricted Access" level, which facilitates the record and monitoring of data requests.

Are any restrictions on data sharing required?

Instruments and intervention resources
These data, which will not contain any results, will be made available as soon as possible. To note, the language of such material will be Portuguese.

Datasets
Dataset files will be kept exclusive for project members until the main manuscript (reporting the intervention results) is accepted for publication. The public data sharing of datasets will be probably made under a "Restricted Access" level at the Zenodo platform. However, data dictionaries and any dataset with only aggregated data will be shared without any restrictions.

Responsibilities and Resources

Who will be responsible for data management?

Our workflow will be initially structured in specific roles and follows:

Basic level

1. Research staff responsible for collecting data

2. Research staff responsible for data entry or simple data processing (e.g., basic routines for data preparation)

Advanced level

3. Researchers responsible for analysing data

4. Data curator

5. Project director

In accordance to the terminology used by the Brazilian General Law for Data Protection (2018), the Principal Investigator (Project director) is the *controller* and other researchers or staff dealing with individual participant data are *operators*.

Reference - General Law for Data Protection (Brasil, 2018):

http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/L13709.htm

What resources will you require to deliver your plan?

Personnel

- four staff members for data handling and data entry: considering undergraduate students with monthly financial aid, the total estimated cost for 1 year is R$ 19.200,00.

- two graduate students for data analyses: considering a student with a standard graduate salary funded by CAPES-Brasil, with a weekly FTE of 0.15 the total estimated cost for 1 year is R$ 7.960,00 (or R$ 3.960,00 per person).

- one postdoctoral fellow for data curation: considering a researcher with a standard postdoctoral salary funded by CAPES-Brasil, with a weekly FTE of 0.1 the total estimated cost for 1 year is R$ 4.800,00.

Hardware

1. Computers with internet access: this infrastructure is available at our laboratory (Hospital de Clínicas de Porto Alegre) and no direct costs apply for software upgrades, maintenance, and other technical services.

2. Cloud storage (primarily, Google Drive): our institution provides a Google Suite account for every project member. No costs apply.

3. Sharing services (OSF and Zenodo): such services are free of costs.

Software

- RedCap: our institution offers this service for every active project, and no direct costs apply.

- Microsoft Excel or Google Sheets: this software is available to the researchers.

- RStudio (free)

- SPPS (licenses already purchased by researchers and by the institution)