

Plan Overview

A Data Management Plan created using DMP Tool

DMP ID: <https://doi.org/10.48321/D1VK6V>

Title: Signature for Pain Recovery IN Teens (SPRINT)

Creator: Emma Biggs - **ORCID:** [0000-0002-0638-6931](https://orcid.org/0000-0002-0638-6931)

Affiliation: Stanford University (stanford.edu)

Principal Investigator: Laura Simons

Data Manager: Alex Tremblay-McGaw

Funder: National Institutes of Health (nih.gov)

Grant: <https://sprint.stanford.edu/>

Template: NIH-Default DMSP

Project abstract:

Up to 5% of teens suffer from high impact chronic musculoskeletal (MSK) pain, affecting all life domains. Discovery of robust markers of the recovery vs. persistence of pain and disability is essential to develop more resourceful and patient-specific treatment strategies and to conceive novel approaches that benefit patients who are refractory to existing treatment options.

The primary goal of this 5-year NIH Multi-Site Mechanistic Study, “**SPRINT**: Signature for Pain Recovery IN Teens,” is to develop a signature comprising of brain, somatosensory, psychological, and immune markers for the recovery of pediatric chronic MSK pain. Specifically, in the *discovery phase* (R61), the aim is to identify biological signatures (e.g., multivariate biological pattern) that will identify adolescents with MSK pain that respond to treatment, and those whose pain persists, as defined by pain severity and functional disability 3 months after baseline. Our expertise in machine learning approaches to extract reliable and prognostic bio-signatures from a large and

complex data set will make this possible. In the *validation phase* (R33), the aim is to validate and clinically evaluate the support tool derived in the first phase in a new cohort of patients. We expect to enroll 150 adolescents in the discovery phase, and 100 in the validation phase.

This project is made possible by the National Institutes of Health HEAL (Helping to End Addiction Long-Term) Initiative, which works to characterize and combat the global opioid crisis.

Start date: 09-28-2020

End date: 02-01-2025

Last modified: 07-08-2024

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

Signature for Pain Recovery IN Teens (SPRINT)

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)

Two cohorts of patients, one for multivariate model discovery (N=150), and the second for independent model validation (N=100), will be recruited from three academic medical centers (Stanford Children's Health Pediatric Pain Management Center, Cincinnati Children's Hospital Medical Center, Toronto Hospital for Sick Children) specializing in pediatric pain management. Data collected will include questionnaires, self-reports, somatosensory responses (quantitative sensory testing; QST), functional and structural magnetic resonance imaging (MRI), and blood-derived immune markers.

Raw questionnaire, self-report and QST data will be collected using a Research Electronic Data Capture (REDCap) system hosted at each site. To maximize accessibility this data will be exported to comma-separated values (.csv) format. Raw MRI data will be collected at each site and managed via Flywheel (hosted at the Lucas Centre, Stanford University). In compliance with IRB restrictions, the raw data collected at SickKids will be de-identified prior to Flywheel upload, while the raw data from Stanford and Cincinnati will be automatically uploaded. Raw data will be collected in DICOM (.dicom) format and converted to the standard Nifti (.nii) format for subsequent processing. Collected blood samples will be regularly transported for storage and processing at Stanford Biobank. Processed data will be made available in .csv format.

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.*

In line with FAIR guidelines, efforts will be made to promote broad and immediate access to data and materials. After complete data collection for each cohort, de-identified Underlying Primary Data and second-level processed data will be deposited in data sharing repositories (see below), along with applicable analysis code.

Underlying Primary Data will be made available to facilitate peer-review and will be publicly available upon request following publication of the main outcomes paper. In line with Stanford policy, access to these data for public use will be subject to acceptance of a Materials Transfer Agreement.

Second-level processed data (available in open source formats) and analysis code will be made publicly available under the Creative Commons Attribution 4.0 Generic License (CC BY 4.0) as per the recommendations of the funding mechanism.

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.

Study metadata will be available at the HEAL central catalog and CEDAR entry, both of which are indexed and searchable resources, and will explicitly include the globally unique and persistent identifier (persistent URL and DOI) of the data repository. Furthermore, the study protocol with detailed information on each measure will be available via Open Access publication and identifiable via DOI.

To ensure the shared data is interpretable and reusable standardized metadata will be provided for each dataset. Specifically, for the Underlying Raw Data and second-level processed data (excluding MRI), shared datasets will consist of .csv files, accompanied by .json dataset descriptions including data dictionaries. MRI data (both Underlying Raw Data and second-level processed data) will be accompanied by .json file descriptions as per the Brain Imaging Data Structure (BIDS) standardized approach.

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.

The Underlying Raw Data (excluding MRI) will be made available in the open source .csv format. The Underlying Raw MRI Data will be made available in the .nii format, which will require specialized software that is freely available. Analysis code used for data processing and analysis will be made available through GitHub in open source languages (R, Python) and SPSS Syntax which is viewable with a standard text editor.

The analysis of the MRI data will use Flywheel gears, with the underlying code available on Github in Python or Bash.

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

Given the multimodal nature of the project, the Interoperable principle of FAIR will be emphasized. The data will be organized by modality, with a standard structure of a <dataset_name>.csv and <dataset_name>_description.json. Study specific file naming standards will be employed to indicate the data modality and stage of processing. Dataset descriptions will include meta-data referring to provenance (i.e., potential processing steps applied), contents (i.e., variable labels), units of measurement, data created, and space for text description (i.e., software used to process data). These

headers will be derived from the DublinCore standard. MRI data will be the exception to this rule, as it will be organized and named according to the BIDS standard.

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; see [Selecting a Data Repository](#))

The primary data repository for data sharing will be the Open Science Framework (OSF), which has been approved by the HEAL initiative as meeting NIH policy for data sharing and management. Within this primary data repository there will be resource specific links to the analysis code used (GitHub) and the MRI data (OpenNeuro), both of which have also been approved by HEAL for this purpose.

For this project we have chosen to use OSF as a primary repository as it allows us to adhere to the FAIR principles while also facilitating the multi-site and multi-modality collaboration within this project. The data will be **Findable** via DOI and persistent URL; meta-data can be publicly and freely **Accessible** and the access to different levels of data can be carefully controlled by the research team to balance institutional and Open Access requirements; multi-modal data can be **Interoperable** within one repository with flexible storage requirements; and finally data will be **Reusable** due to the clear data usage licensing and ability to link rich metadata with clear provenance information.

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.

- DOI and persistent URL for data repository
- ORCID iDs for research team members (linked to data repository and publications)
- DOI for study protocol paper and other future output
- Publications and study descriptions will include relevant MeSH terms to make these resources easily searchable.

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.

Data will be made available upon request at the time of publication of the main outcomes paper (per cohort) and will be preserved indefinitely.

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. See [Frequently Asked Questions](#) for examples of justifiable reasons for limiting sharing of data.

Access and reuse of the resulting scientific data will be limited and approved by a committee of the SPRINT (co-)PIs. If access is granted by the committee then both parties will also need to reach a Data Usage Agreement which will be approved by Stanford University (as lead institution) and the primary institution of the recipient.

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).

There will be a publication charter and committee in place to review requests for data access to ensure scientific integrity of studies conducted using these data, and to avoid overlap in ancillary studies using these data. As per Stanford policy, a Materials Transfer Agreement / Data Usage Agreement will also be reached prior to sharing of materials or data.

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).

To protect participant privacy, all shared data will be de-identified. Furthermore, participants will be asked for explicit permission to share data, if this is not granted then data will not be deposited to the data repository for potential sharing. The primary methods for de-identification will be the removal of names and dates (birth, study visit, pain onset, etc) and skull stripping of MRI data.

In line with the IRB agreement of SickKids (Toronto, Canada), MRI data will be de-identified prior to sharing with US-based collaborators.

Prior to data sharing, data use agreements will be put in place to assess on a case by case basis any additional limitations that may be needed.

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 (Laura Simons), will be responsible for the day-to-day oversight of 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 data stewardship. Oversight for data collection, quality assurance, processing, and curation will be undertaken by each modality lead.
