ONTARIO BRAIN INSTITUTE | INSTITUT ONTARIEN DU CERVEAU

OBI Data Quality Framework

Stage 2 – Data Quality Reporting Guide

Version: 1.2 - Date: August 8, 2024

Note: this document and the procedures herein are designed for use with active studies that are collecting data and have not had their final data release. The outline reporting procedures may be subject to change.

1. Introduction

OBI aims to ensure high-quality data is made available for collaboration and sharing and achieves accuracy, reliability, timeliness, completeness, precision, and integrity. To validate that data is high quality, the OBI Data Quality team will work with Data Producers (DP) to generate the following documentation (across all data modalities) on a quarterly basis:

- Study Quality Checklist: Stage 2
- Participant status
- *Missing data (currently dependent on modality and upon discussion with OBI)*

This documentation will be created using data from multiple sources including automated data exports from electronic data capture (EDC) tools (e.g., REDCap API export), data transfers, and quality analysis procedures (e.g., REDCap Data Quality Rules and Data Workflow Resolution, DPs' quality assurance procedures), as appropriate. The OBI Data Quality team will review the documentation within the data capture tools or given by the DP and will conduct quarterly QA/QC checks on the available data. Upon completion, the OBI Data Quality team will generate a data quality report analyzing data missingness and formatting, as applicable.

Throughout this process, data will be organized and analyzed at three levels as described below. Please refer to document "OBI Data Quality Glossary of Terms and List of Acronyms" for additional definitions.

- Level 1
    - Level 1 applies to **data packages** per modality (e.g., clinical, imaging, genomics, wearables). Packages ***must*** include data files, metadata files, and documentation/data provenance files.
- Level 2

- Level 2 applies to **data records**. A record is data from one instrument for one participant at one timepoint (e.g., the WHOQOL-BREF or a NIfTI file of a T1-weighted MRI scan for one participant and may include multiple scans for different tasks/experiments).

- Level 3
  - Level 3 applies to **data items**. For tabular data, an item is an individual value in a record (e.g., question #5 on the WHOQOL-BREF instrument for one participant). For imaging data, an item is data from one scan for one participant for one task at one timepoint (e.g., an individual T1-weighted DICOM file from one participant).

2. Missing Data Coding Guide

The following table outlines and defines the codes used to flag missing data. In an effort to maximize the trustworthiness of the data and support standardization across research programs, these codes offer sufficient specificity addressing various situations that can be present across data levels, modalities, and studies. Please see the Data Producer Requirements section for more details on their use within each modality.

Note:

- *For the following table, "assessment" refers to a clinical instrument, task, biological sample, or neuroimaging scan.*
- *Not all missing codes are applicable to all studies, data levels, modalities, or stages of data collection/curation. Only applicable missing codes should be included in REDCap projects (see Stage 1 document "OBI DQF REDCap Tools & Procedures" for the list of relevant codes).*
- *Missing data codes are to be used when data is expected but missing. For example, if assessments are not administered due to age according to protocol, this is not considered missing.*
- *The OBI Data Quality team will provide a list of examples for each code that will be updated as needed (see "OBI DQF Missing Data Codes – Examples"). Should there be questions about which code is relevant for a particular situation, please contact the OBI Data Quality team.*

Code	Definition
unable_to_perform	The participant could not participate in the assessment/task due to a physical or cognitive condition.
task_declined	The participant chose not to participate in the assessment/task or declined to provide a response.
out_of_time	The participant was unable to complete the assessment/task within the scheduled timeslot.
admin_error	The administrator of the assessment/task made a mistake that caused the entry to be invalid or otherwise unobtainable. E.g., miscalibration of equipment by administrators; incorrect administration; scheduling conflicts.
technical_error	The equipment or tools (hardware or software) used for data acquisition, collection, or processing failed. E.g., computer crashed; power failure; response button box or keyboard was not properly connected or defective.
missing_other	Data is missing for a reason that is not explained by any of the available missing data codes. ***Please use sparingly and provide a description if possible (a comment in REDCap for clinical data if using the Data Quality tools or text box in the Missing Data Flagging form).***
value_unknown	The participant could not provide a response to a question because they did not know how to answer it. ***Clinical only. To be used if the field does not include a "do not know" option.*** *
not_applicable	The assessment/task was not performed, or a value could not be derived because it did not apply to the participant. ***This code should only be used for Level 3 data.*** *
artifacts_present	Data are not usable due to artifacts. E.g., unusable FLAIR because of foreign bodies or motion. ***Imaging only.***

*Please note these scenarios are not considered missing data, as "do not know" is an answer to a given question and any fields that are "not applicable" are not expected to contain data. For simplicity, these are included in the Missing Data Coding Guide so that this information can be captured in one place, and all fields can have a data point or label.

3. Data Producer Requirements

In addition to the minimally curated data, DPs are required to provide data quality files as part of the Data Quality Reporting process and a completed "Stage 2" section of the Study Quality Checklist provided in Stage 1 (document #5). These data quality files can be exported by the OBI Data Quality team depending on the method of collection for this data. If DPs use their own methods of documenting the required information, the OBI Data Quality team provides

reporting templates in the Templates folder ("Participant Status Template.xlsx", "Missing Data Template.xlsx"). Please use these templates as necessary (or other agreed-upon formats) to provide the required information once minimal curation on the data is complete for each quarterly (or as agreed upon with OBI) Data Quality Reporting iteration.

***Note**: for any data that will be transferred to Brain-CODE, please provide a completed "OBI Data Transfer Plan Template" (as described in the Stage 2 main document) **prior** to the initial data transfer.*

3.1 Study Quality Checklist: Stage 2

Please provide a completed Study Quality Checklist for the section on Stage 2.

- See Stage 1 document #5 "05. OBI DQF Study Quality Checklist"

3.2 Participant Status

If the Participant Status form is being used as described in the Stage 2 main document, the OBI Data Quality team will review this REDCap form as part of the Data Quality Reporting process (see Stage 1 training document "OBI DQF REDCap Tools & Procedures" to see the design and use of the form). If the form is not used, the DP is expected to provide the same information for all participants using any current procedures or using the template provided by OBI (see Stage 2, "Participant Status Template.xlsx"):

- Whether the participant is active, has completed the study, has withdrawn, or was lost to follow-up
- If withdrawn:
  - The reason for withdrawal
  - The date of withdrawal
- Which timepoints the participant had completed

3.3 Missing Data

The provided missing data codes should be used at least on Level 2 data. Where possible, DPs are asked to use applicable codes on Level 3 data (see Stage 2, "Missing Data Template").

Clinical

- Level 3 data (items): With the use of REDCap's missing data codes feature and Data Quality tools, missing items will be identified and analyzed (see Stage 1 document "OBI DQF REDCap Tools & Procedures"). These codes are only to be used within partially completed records, not entirely blank or missing ones. The Data Quality tools output will be reviewed by the OBI Data Quality team. If REDCap is not being used, further discussion will be had with OBI on how to best identify missing data and provide documentation.
- Level 2 data (forms): The OBI Data Quality team has developed the Missing Data Flagging form to identify missing forms within REDCap projects and assign them missing data codes (see Stage 1 document "OBI DQF REDCap Tools & Procedures"). If REDCap is not being used, further discussion will be had with OBI on how to best identify missing data and provide documentation.

Imaging, Genomics, etc.

- Further discussion will be had with OBI on how to best identify missing data and provide documentation.