RADx Data Hub Global Instruction Manual

Introduction

This file is the NIH RADx Data Hub Global Codebook Instruction Manual and contains the RADx Data Hub Global Codebook Orientation Guide, RADx Global Codebook Mapping Logic Overview, and RADx Global Codebook Transformation Guide. This document is the supplemental instruction guide to the RADx Global Codebook and includes an overview of the RADx Global Codebook, an overview of the mapping logic process, and instruction set for harmonizing the 12 NIH Required Common Data Elements (CDEs) across the four (4) RADx programs: RADx-UP, RADx-rad, RADx-TECH, and Digital Health Technologies (DHT) at each Data Coordination Center (DCC).

1.0 RADx Data Hub Global Codebook Orientation Guide

1.1 Introduction

The NIH RADx Data Hub Global Codebook contains the detailed cross-program mapping logic for the 12 NIH RADx Required Common Data Elements (CDEs). This harmonization effort includes mapping for all 12 NIH Required CDEs collected by four (4) RADx programs: RADx-UP, RADx-rad, RADx-TECH, and Digital Health Technologies (DHT) at each Data Coordination Center (DCC).

1.2 RADx Global Codebook Contents

Sheet 2: "2_RADx_Global_Codebook" contains the following:

- "CDE Names" (column A): The NIH Required 12 CDE names
- "RADx Global Prompt" (column B): All harmonized survey prompts
- "Variable name" (Column C): The variable name (column header) for each data point
- "Responses" (Column D): Accepted response type or codified response options for each survey prompt with associated integer identifiers

Sheet 3-6: Global Codebook compared to each DCC's Data Dictionary

Each DCC has a separate sheet containing their respective data dictionaries compared against the RADx Global Codebook. The RADx Global Codebook entries (Columns B-D) and the respective DCC-specific columns (Columns E-G) are aligned to the appropriate CDEs. If multiple DCC-specific variables map together within a single Global variable, the DCC-specific variables are aligned in the Multi-Mapped Variable Columns (beginning in Column H).

Sheets 7-24: Individual CDE Mapping

Each CDE (Identity, Race, Ethnicity, Age, Sex, Education, Domicile, Employment, Insurance Status, Disability Status, Medical History, Symptoms, and Health Status) is separated into sheets in which the CDEs are compared across the four (4) programs.

These sheets contain:

- RADx Program (Column A): RADx Global, RADx-UP, RADx-rad, RADx-TECH, or DHT (Digital Health)
- Prompt (Column B): The prompts for each CDE provided by each program
- Variable (Column C): The variable for each of the prompts provided by each program
- Responses (Column D): Accepted response type or codified response options for each survey prompt with associated integer identifiers

Additionally, these sheets contain mapping strategies (bottom) that display how the program-specific variables and responses are harmonized into the Global variables and responses. The Global prompts reflect language that is inclusive to each harmonized program-specific prompt.

The Disability, Medical History, Symptoms, and Health Status CDEs are further expanded because they do not have a one-to-one relationship with their respective variables and prompts. For example, the Disability CDE has prompts associated with general disability, deafness, blindness, memory, and mobility-each with their unique variables and responses. These prompts, variables, and responses also have been mapped across the four (4) programs.

2.0 RADx Global Codebook Mapping Logic Overview

2.1 Introduction

Utilizing the NIH RADx Required 12 Common Data Elements (CDE) and guidance, the RADx Global CDEs Codebook was developed to inform data harmonization and mapping for the Required CDE data submitted to the RADx Data Hub.

- **Phase 1:** Developed first version of the RADx Global CDE Codebook that included all values, variables, and accepted ranges
- **Phase 2:** Reviewed each RADx DCC data dictionary/codebook at the prompt and response level to determine the Required 12 CDEs adherence
- **Phase 3:** Addressed DCC response level inconsistencies to inform response options (when applicable)
    - DCC #1 collects "Male", "Female", "Prefer not to answer"
    - DCC #2 collects "Male", "Female", "Intersex"
- **Phase 4:** Mapped response values across all DCCs to the RADx Global CDE Codebook

DCC #1	DCC #2	DCC #3	RADx Codebook
1, Female | 1 → 0	**2,** Female | 2 → 0	**2,** Female | 2 → 0	**0,** Female
0, Male	**1,** Male	**1,** Male	**1,** Male
	3, Intersex		**2,** Intersex
2, Non-binary			**3,** Non-binary
96, None describe me	**4,** None describe me		**98,** None describe me
99, Prefer not to answer		**88,** Prefer not to answer	**99,** Prefer not answer

3.0 RADx Global Codebook Transformation Guide

3.1 Introduction

In this document, "transformation" is defined as the conversion of the 12 NIH Required CDEs as collected by each DCC to the harmonized CDEs defined in the RADx Global Codebook. The transformation will result in converting 1) variable names, 2) response options, and 3) the associated codified values for the 12 NIH Required CDEs to a single, harmonized set.

3.2 Transformation Guidance

This section provides an overview and step-by-step instruction set detailing the process for transformation. Transformation of data is the responsibility of each RADx DCC and will result in two files per study: the raw data file (pre-harmonization) and the transformed data file (post-harmonization), which must both be submitted to the RADx Data Hub. Both files will be retained within the RADx Data Hub, to ensure that the detailed DCC-specific responses are available for analysis in addition to the harmonized responses. The proper naming convention is required to differentiate between the original and the transformed dataset.

Below is a list of instructions for transforming the data files before submission into the RADx Data Hub.

- **STEP 1:** Ensure that the raw data file has undergone the de-identification process in accordance with HIPPA guidelines.
- **STEP 2:** Ensure that the data file adheres to the respective RADx DCC data dictionary for all variable names, response options, and associated codified values.
- **STEP 3:** Identify the 12 NIH Required CDEs (Tier 1) within the data file that will undergo transformation in accordance with the RADx Global Codebook. *Note: Harmonization only applies to the 12 NIH Required CDEs.*
- **STEP 4:** Consult with the RADx Global Codebook sheets 7-24 for the appropriate DCC's CDE mapping logic.
- **STEP 5:** Create a new file that contains the transformed codified values for the 12 NIH Required CDEs (Tier 1). There will now be two (2) distinct files, a transformed file, and a raw file.
    a. The transformed file contains:
        i. Transformed Tier 1 CDEs: Transformed files must have all Tier 1 CDEs moved to the leftmost columns within the data file.
        ii. Non-transformed Tier 2 data elements.
    b. The raw file contains:
        i. Non-transformed Tier 1 CDEs.
        ii. Non-transformed Tier 2 data elements.
    c. Utilize the RADx Data Hub file naming convention as detailed below:
        i. Guidance: For the Original Data File, please add "_origincopy" to the end of the file name. For the Transformed Data File, please add "_transformcopy" to the end of the file name.
        ii. Example:
            1. Original Data File Name: *filename123_origincopy.csv*
            2. Transformed Data File Name: *filename123_transformcopy.csv*

3.3 Transformation Example: RADx-TECH Ethnicity CDE Mapping

- RADx-TECH collects more response options for Ethnicity CDE than other RADx programs.
- Response options are merged as a broader "Yes, of Hispanic or Latino origin" response.
  - The DCC program officer must transform "1" "2" or "3" values into "1" for harmonization.
- Original RADx-TECH responses are retained in DCC-specific data sets to preserve data complexity.

RADx-TECH Response	RADx-TECH Transform	RADx Global Codebook
ethnicity	nih_ethnicity	nih_ethnicity
0,No	0 → 0	0, No, not of Hispanic or Latino
1,Yes: Mexican, Mexican American Or Chicano;2,Yes: Puerto Rican;3,Yes: Cuban	1 → 1 2 → 1 3 → 1	1, Yes, of Hispanic or Latino origin
99,Yes: Other Or Mixed Hispanic, Latino Or Spanish Origin	99 → 97	97, Other
77,Don't Know	77 → 98	98, Don't Know
88,Prefer Not To Answer	88 → 99	99, Prefer not to answer

Pre-harmonization
Responses retained in RADx Data Hub as "RADx-TECH original dataset"

Post-harmonization
Responses retained in RADx Data Hub as "RADx-TECH transformed dataset"

*Four (4) CDEs have more than one (1) variable associated with it (Disability Status, Medical History, Symptoms, and Health Status). Transformation must occur on every variable and response option within these CDEs.

3.4 NIH Required CDE Exemption & Exclusion Mapping Process

Studies performed only on participants defined by a specific CDE characteristic, may not explicitly collect that CDE in the datasets

Studies with CDE related inclusion criteria are required to include all CDEs in their datasets, therefore the PI should add the CDE columns with appropriate variable and codified values from the RADx Global Codebook

Example: A study enrolls only female participants and has an exemption for the Employment CDE. RADx-rad study only enrolls females as participants. The Sex CDE may be excluded from collected data, but the Principal Investigator must still include a column with the variable "nih_sex" and values of "0" for all rows

"Sex" CDE Not Collected Based on Study Inclusion Criteria

1. DCC identifies CDEs present in the study description
2. Locate variable name in RADx Global Codebook
3. Add column in study data file:
   a. Add RADx Global Codebook CDE variable
   b. Enter appropriate variable code in each row

1. Identify missing CDE

CDE Name
Sex

2. Review RADx Global Codebook

nih_sex
0, Female
1, Male
2, Intersex
3, Non-binary
98, None describe me
99, Prefer not answer
101, Not Collected

3. Add column including variable and value

nih_sex
0