Data Dictionary for Documents\_Datasets\_Data

**Investigators:** Paije Wilson (MLIS) and Vojtech Huser (MD, PhD)

**Contact:** For questions, please contact the PI (Paije Wilson) at [paije.wilson@wisc.edu](mailto:paije.wilson@wisc.edu)

**Note:** For additional information relating to the study, refer to the full study report.

**Additional Note:** Instances of "PW" in this data dictionary refer to the PI, Paije Wilson

Contents

[**1. Information Exported from ClinicalTrials.gov (CTG)** 2](#_Toc118205101)

[A. Location in Excel Sheet 2](#_Toc118205102)

[B. Data Elements 2](#_Toc118205103)

[**2. Study Documents Data** 2](#_Toc118205104)

[A. Location in Excel Sheet 2](#_Toc118205105)

[B. Data Elements 2](#_Toc118205106)

[**3. Datasets Data** 3](#_Toc118205107)

[A. Location in Excel Sheet 3](#_Toc118205108)

[B. Data Elements 3](#_Toc118205109)

# **1. Information Exported from ClinicalTrials.gov (CTG)**

**Note:** Exported directly from ClinicalTrials.gov (CTG) on 6/6/2022.

## **A. Location in Excel Sheet**

Columns A through AA

## **B. Data Elements**

Refer to CTG's [Glossary of Common Site Terms](https://www.clinicaltrials.gov/ct2/about-studies/glossary) for definitions.

# **2. Study Documents Data**

## **A. Location in Excel Sheet**

Columns AB through AG

## **B. Data Elements**

**Protocol (*Column AB*)** (Is there a protocol link in the Study Documents section of the CTG record?)

* **Y: Yes** (A protocol link is present in the Study Documents section of the CTG record)
* **N: No** (No protocol is linked in the Study Documents section of the CTG record)

**Protocol link functional (*Column AC*)** (Is the protocol link in the CTG record functional?)

* **Y: Yes** (The protocol link in the CTG record is functional)
* **N: No** (The protocol link in the CTG record is not functional)
* **I: Irrelevant** (There is no link to a protocol in the CTG record)

**Informed consent form (*Column AD*)** (Is there an informed consent form link in the Study Documents section of the CTG record?)

* **Y: Yes** (An informed consent form link is present in the Study Documents section of the CTG record)
* **N: No** (No informed consent form link is present in the Study Documents section of the CTG record)

**Informed consent form link functional (*Column AE*)** (Is the informed consent form link in the CTG record functional?)

* **Y: Yes** (The informed consent form link in the CTG record is functional)
* **N: No** (The informed consent form link in the CTG record is not functional
* **I: Irrelevant** (There is no link to an informed consent form in the CTG record)

**Statistical analysis plan (*Column AF*)** (Is there statistical analysis plan link in the Study Documents section of the CTG record?)

* **Y: Yes** (A statistical analysis plan link is present in the Study Documents section of the CTG record)
* **N: No**: (No statistical analysis plan link is present in the Study Documents section of the CTG record)

**Statistical analysis plan link functional (*Column AG*)** (Is the statistical analysis plan link in the CTG record functional?)

* **Y: Yes** (The statistical analysis plan link in the CTG record is functional)
* **N: No** (The statistical analysis plan link in the CTG record is not functional)
* **I: Irrelevant** (There is no link to a statistical analysis plan in the CTG record)

# **3. Datasets Data**

## **A. Location in Excel Sheet**

Columns AH through AR

## **B. Data Elements**

**Plan to share data (*Column AH*)** (Do the researchers state they plan to share their data in their "Plan to Share IPD" in the "Individual Participant Data (IPD) Sharing Statement" section of the CTG record?)

* **Y: Yes** (The researchers state "yes" in their "Plan to Share IPD" in the CTG record)
* **N: No** (The researchers state "no" in their "Plan to Share IPD" in the CTG record)
* **U: Undecided** (The researchers state "undecided" in their "Plan to Share IPD" in the CTG record)
* **I: Irrelevant** (There is no "Individual Participant Data (IPD) Sharing Statement" in the CTG record)

**Plan description/access criteria (*Column AI*)** (Do the researchers state they plan to share their data with non-affiliates in either their "Plan Description" or their "Access Criteria" in the "Individual Participant Data (IPD) Sharing Statement" section of the CTG record?)

**Note:** Researchers must state they plan to share their IPD, either identified or deidentified, in either the "Plan Description" or "Access Criteria" in order to be categorized as a "yes." Plans to only share summary data or genetic sequencing data will be categorized as a "no." Statements that IPD will be only shared within the research team/affiliated institution(s) will also result in a "no" categorization.

* **Y: Yes** (The researchers state they plan to share their data in the "Plan Description" and/or "Access Criteria" in the CTG record)
* **N: No** (The researchers state they do not plan to share their data in the "Plan Description" or "Access Criteria" in the CTG record)
* **U: Undecided** (The researchers state they are undecided on whether to share their data in the "Plan Description" or "Access Criteria" in the CTG record)
* **I: Irrelevant** (There is no "Plan Description" nor "Access Criteria" in the CTG record)

**Plan inconsistencies (*Column AJ*)** (Is there an inconsistency between the "Plan to Share IPD" and "Plan Description"/"Access Criteria" in the CTG record? E.g., "No" for the "Plan to Share IPD," "Yes" for the "Plan Description")

* **Y: Yes** (There are inconsistencies between the "Plan to Share IPD" and "Plan Description"/"Access Criteria" in the CTG record)
* **N: No** (There are no inconsistencies between the "Plan to Share IPD" and "Plan Description"/"Access Criteria" in the CTG record)
* **I: Irrelevant** (Either:
  + 1) there is no "Individual Participant Data (IPD) Sharing Statement" section in the CTG record, or
  + 2) there is a "Plan to Share IPD" but no "Plan Description" nor "Access Criteria" in the CTG record)

**Timeframe (*Column AK*)** (What is the timeframe for sharing data listed in the "Time Frame" in the "Individual Participant Data (IPD) Sharing Statement" section of the CTG record?)

* **[Textual entries for specified timeframes]**
* **U: Unspecified** (The researchers state "yes" in their "Plan to Share IPD" and/or "Plan Description"/"Access Criteria," but do not include a timeframe for sharing their data in the "Time Frame" in the CTG record)
* **I: Irrelevant** (Either:
  + 1) there is no "Individual Participant Data (IPD) Sharing Statement" in the CTG record, or
  + 2) the researchers state "no" or "undecided" in their "Plan to Share IPD" and in their "Plan Description"/"Access Criteria" in the CTG record, or
  + 3) the researchers state "no" or "undecided" in their "Plan to Share IPD" and don't have a "Plan Description" nor "Access Criteria" in the CTG record)

**Mechanism for getting access to data (*Column AL*)** (What is the mechanism for getting access to the data listed in either the "Plan Description" or the "Access Criteria" in the "Individual Participant Data (IPD) Sharing Statement" section of the CTG record?)

Mechanisms for accessing summary data or genetic sequencing data will not be included (i.e., only mechanisms for accessing IPD are included). Mechanisms by which *only* the research team(s) or affiliated institution(s) can access data will also be excluded (but will be mentioned in the notes section) (i.e., only mechanisms for non-affiliates accessing data will be included).

* **P: Data sharing platform** (The "Plan Description" and/or the "Access Criteria" states requestors may access the data via a data sharing platform)
* **R: Researcher contact** (The "Plan Description" and/or the "Access Criteria" states requestors must contact the PI/researcher to get access to the data. This also includes statements of "data can be accessed upon request" when a platform isn't mentioned)
* **C: Could not determine** (The CTG record states "yes" in their "Plan to Share IPD" and/or in their "Plan Description"/"Access Criteria", but the CTG record does not mention the mechanism for getting access to the data)
* **I: Irrelevant** (Either:
  + 1)there is no "Individual Participant Data (IPD) Sharing Statement" section in the CTG record, or
  + 2) the researchers state "no" or "undecided" in their "Plan to Share IPD" and in their "Plan Description"/"Access Criteria" in the CTG record, or
  + 3) the researchers state "no" or "undecided" in their "Plan to Share IPD" and have no "Plan Description" nor "Access Criteria" in the CTG record)

**Email address provided (*Column AM*)** (In cases where researcher contact is listed as the mechanism for getting access to data in either the "Plan Description" or the "Access Criteria" in the "Individual Participant Data (IPD) Sharing Statement" section of the CTG record, is an email address provided anywhere in the CTG record?)

* **Y: Yes** (An email address is provided in the CTG record)
* **N: No** (An email address is not provided in the CTG record)
* **I: Irrelevant** (Either:
  + 1) there is no "Individual Participant Data (IPD) Sharing Statement" section in the CTG record, or
  + 2) researcher contact is not listed as a mechanism for getting access to the data in the "Plan Description" nor the "Access Criteria" in the CTG record)

**Platform name provided (*Column AN*)** (Is the name of the data sharing platform in which the researchers plan to share their data listed in the either the "Plan Description" or the "Access Criteria" in the "Individual Participant Data (IPD) Sharing Statement" section of the CTG record?)

* **Y: Yes** (The name of the data sharing platform in which the researchers plan to share their data is listed in the "Plan Description" and/or the "Access Criteria" in the CTG record)
* **N: No** (The CTG record states requestors may access the data via a data sharing platform in the "Plan Description" and/or "Access Criteria," but neither the "Plan Description" nor the "Access Criteria" list the name of the data sharing platform in which the researchers plan to share their data)
* **I: Irrelevant** (Either:
  + 1) there is no "Individual Participant Data (IPD) Sharing Statement" section in the CTG record, or
  + 2) data sharing platform is not listed as a mechanism for getting access to the data in the "Plan Description" nor the "Access Criteria" in the CTG record)

**Planned Platform (*Column AO*)** (What is the name of the data sharing platform in which the researchers plan to share their data?)

* **[Name of platform]**
* **I: Irrelevant** (Either:
  + 1) there is no "Individual Participant Data (IPD) Sharing Statement" section in the CTG record, or
  + 2) data sharing platform is not listed as a mechanism for getting access to the data in the "Plan Description" nor the "Access Criteria" in the CTG record

**Discoverability (*Column AP*)** (For datasets shared on a data sharing platform, how discoverable is the trial's associated dataset record?)

**Note:** For trials that named a specific data sharing platform as a mechanism for sharing data in either the "Plan Description" or the "Access Criteria" in the "Individual Participant Data (IPD) Sharing Statement" section of the CTG record, but didn't provide a direct link to the dataset record, PW searched the platform for the dataset record using the trial's NCT number or, if the latter retrieved no results, searched the platform using the trial's title in its corresponding CTG record

* **D: Direct** (There is a direct, functional link to the dataset record present in the "Individual Participant Data (IPD) Sharing Statement" section of the CTG record)
* **A: Additional searching required** (PW had to search the data sharing platform mentioned in the "Plan Description" and/or the "Access Criteria" of the CTG record to locate the dataset record)
* **N: Not discovered** (A specific data sharing platform was mentioned in the "Plan Description" and/or the "Access Criteria" in the CTG record, but PW was unable to locate the dataset record in the platform)
* **P: Platform is private or otherwise inaccessible** (A specific data sharing platform was mentioned in the "Plan Description" and/or the "Access Criteria" in the CTG record, but PW couldn't search for the dataset record because the platform was inaccessible)
* **I: Irrelevant** (Either:
  + 1) there is no "Individual Participant Data (IPD) Sharing Statement" section in the CTG record, or
  + 2) data sharing platform is not listed as a mechanism for getting access to the data in the "Plan Description" nor the "Access Criteria" in the CTG record, or
  + 3) no platform name was mentioned and no link to the dataset record was present in the "Plan Description" nor the "Access Criteria" in the CTG record)

**Link back to CTG (*Column AQ*)** (For CTG records were an associated dataset was found in a data sharing platform, is there a link from the dataset record back to the CTG record?)

* **Y: Yes** (There is a link back to the CTG record present in the dataset record)
* **N: No** (There is no link back to the CTG record present in the dataset record)
* **I: Irrelevant** (The dataset record is unavailable or could not be located)

**Notes (*Column AR*)** (Notes that informed data entry for Datasets Data)

* **[Textual entries]**