

Data Overview Sheet — Clinical Trial Success & Dropout Analysis

Field Name	Description / Meaning	Example	Data Type	Notes
NCT Number	Unique identifier assigned to each clinical trial on ClinicalTrials.gov	NCT04567891	Text	Primary key; used to merge datasets
Study Title	Full title describing the study's purpose	<i>A Study to Evaluate Drug X in Type 2 Diabetes</i>	Text	Useful for referencing study context
Study Status	Current progress stage of the trial	Completed	Text	Key target variable for success/dropout classification
Study Results	Whether outcome results have been posted	With Results	Text	Helps assess reporting transparency
Conditions	Health condition(s) studied	Type 2 Diabetes Mellitus	Text	Use for grouping or filtering trials
Interventions	Drug, procedure, or behavioral treatment tested	Drug: Metformin	Text	May contain multiple entries separated by semicolons
Primary Outcome Measures	Main measured endpoints	Change in HbA1c levels after 24 weeks	Text	Unstructured text; consider text-mining
Sponsor	Organization leading the trial	Pfizer Inc.	Text	Combine with “Funder Type” for sponsor impact analysis
Collaborators	Supporting institutions	Harvard Medical School	Text	Often missing or sparse
Sex	Participant sex eligibility	All	Categorical	Values: Male / Female / All
Age	Age range eligible for participation	18 Years to 65 Years	Text	Eligibility filter; not actual ages
Phases	Stage of drug testing	Phase 3	Categorical	Values: Phase 1–4, N/A
Enrollment	Number of participants enrolled	350	Numeric	Can indicate study scale; use in dropout analysis
Funder Type	Type of study sponsor or funding source	Industry	Categorical	Values: Industry / NIH / Other / U.S. Fed Agency
Study Type	Study design classification	Interventional	Categorical	Key for separating experimental vs. observational data
Start Date	Study initiation date	Jan 15, 2018	Date	Useful for time trend analysis
Completion Date	Study end date	Mar 20, 2022	Date	Compare with Start Date for duration metrics
Locations	Sites where trial conducted	Johns Hopkins Hospital, Maryland	Text	Multi-location possible; clean for mapping

General Field Notes:

- Data Source: *ClinicalTrials.gov* (publicly available registry)
- File Name: clinical_trials_raw_data.xlsx
- Total Columns: 18
- Each row = one clinical trial record
- Missing data (especially in Collaborators, Study Results) is common — handle carefully during analysis
- Key analyses: trial success prediction, dropout trend identification, sponsor-wise performance