

# Package ‘r4subdata’

February 26, 2026

**Title** Example Datasets for Clinical Submission Readiness

**Version** 0.1.1

**Description** Provides realistic synthetic example datasets for the R4SUB (R for Regulatory Submission) ecosystem. Includes a pharma study evidence table, ADaM (Analysis Data Model) and SDTM (Study Data Tabulation Model) metadata following CDISC (Clinical Data Interchange Standards Consortium) conventions (<<https://www.cdisc.org>>), traceability mappings, a risk register based on ICH (International Council for Harmonisation) Q9 quality risk management principles (<<https://www.ich.org/page/quality-guidelines>>), and regulatory indicator definitions. Designed for demos, vignettes, and package testing.

**License** MIT + file LICENSE

**URL** <https://github.com/R4SUB/r4subdata>

**BugReports** <https://github.com/R4SUB/r4subdata/issues>

**Depends** R (>= 4.2)

**Imports** tibble

**Suggests** testthat (>= 3.0.0)

**Config/testthat/edition** 3

**Encoding** UTF-8

**RoxygenNote** 7.3.3

**LazyData** true

**NeedsCompilation** no

**Author** Pawan Rama Mali [aut, cre, cph]

**Maintainer** Pawan Rama Mali <prm@outlook.in>

**Repository** CRAN

**Date/Publication** 2026-02-26 13:20:02 UTC

## Contents

adam_metadata . . . . .	2
dataset_dictionary . . . . .	3
evidence_pharma . . . . .	3
list_datasets . . . . .	4
regulatory_indicators . . . . .	5
risk_register_pharma . . . . .	5
sdtm_metadata . . . . .	6
trace_mapping . . . . .	7

<b>Index</b>	<b>9</b>
--------------	----------

---

adam_metadata	<i>ADaM Variable-Level Metadata</i>
---------------	-------------------------------------

---

### Description

ADaM (Analysis Data Model) variable-level metadata for ADSL (Subject-Level Analysis Dataset, 16 vars), ADAE (Adverse Events Analysis Dataset, 10 vars), and ADLB (Laboratory Results Analysis Dataset, 10 vars). Follows CDISC (Clinical Data Interchange Standards Consortium) ADaM conventions.

### Usage

```
adam_metadata
```

### Format

A tibble with 36 rows and 6 columns:

**dataset** Character. ADaM dataset name (ADSL, ADAE, ADLB).

**variable** Character. Variable name.

**label** Character. Variable label.

**type** Character. Variable type (Char or Num).

**length** Integer. Variable length.

**format** Character. SAS (Statistical Analysis System) format (or NA).

### Source

Synthetic metadata based on CDISC ADaM (Analysis Data Model) standards.

### Examples

```
data(adam_metadata)
table(adam_metadata$dataset)
```

---

dataset_dictionary	<i>Dataset Column Dictionary</i>
--------------------	----------------------------------

---

**Description**

Returns column names, types, and descriptions for a given r4subdata dataset.

**Usage**

```
dataset_dictionary(dataset)
```

**Arguments**

dataset            Character. Name of the dataset (e.g., "evidence\_pharma").

**Value**

A tibble with columns: column, type, description.

**Examples**

```
dataset_dictionary("evidence_pharma")  
dataset_dictionary("adam_metadata")
```

---

evidence_pharma	<i>Pharma Study Evidence Table</i>
-----------------	------------------------------------

---

**Description**

A realistic evidence table for study CDISCPILLOT01 (Clinical Data Interchange Standards Consortium Pilot Study 01) covering all four R4SUB (R for Regulatory Submission) pillars (quality, trace, risk, usability) with 250 rows and 18 indicators across multiple datasets and sources.

**Usage**

```
evidence_pharma
```

**Format**

A tibble with 250 rows and 17 columns:

**run\_id** Character. Unique run identifier.

**study\_id** Character. Study identifier (CDISCPILLOT01).

**asset\_type** Character. Asset type: dataset, define, program, validation, spec, other.

**asset\_id** Character. Asset identifier (e.g., ADSL, define.xml).

**source\_name** Character. Source of the evidence (e.g., pinnacle21).  
**source\_version** Character. Version of the source tool.  
**indicator\_id** Character. Indicator identifier (e.g., Q-MISS-VAR).  
**indicator\_name** Character. Human-readable indicator name.  
**indicator\_domain** Character. Domain: quality, trace, risk, usability.  
**severity** Character. Severity: info, low, medium, high, critical.  
**result** Character. Result: pass, fail, warn, na.  
**metric\_value** Numeric. Metric value (if applicable).  
**metric\_unit** Character. Unit for metric\_value.  
**message** Character. Descriptive message.  
**location** Character. Location reference (e.g., ADSL:AGE).  
**evidence\_payload** Character. JSON payload with additional details.  
**created\_at** POSIXct. Timestamp when evidence was created.

### Source

Synthetic data based on the CDISC (Clinical Data Interchange Standards Consortium) Pilot Study 01 structure.

### Examples

```
data(evidence_pharma)
head(evidence_pharma)
table(evidence_pharma$indicator_domain)
```

---

list_datasets	<i>List Available r4subdata Datasets</i>
---------------	--

---

### Description

Returns a summary of all datasets included in the r4subdata package.

### Usage

```
list_datasets()
```

### Value

A tibble with columns: name, description, n\_rows, n\_cols.

### Examples

```
list_datasets()
```

---

regulatory\_indicators *Regulatory Indicator Definitions*

---

### Description

Reference table of 30 indicator definitions across all four R4SUB (R for Regulatory Submission) domains (quality, trace, risk, usability). Each indicator has a unique ID, default severity, typical source, and descriptive tags.

### Usage

```
regulatory_indicators
```

### Format

A tibble with 30 rows and 7 columns:

**indicator\_id** Character. Unique indicator identifier.

**indicator\_name** Character. Human-readable indicator name.

**domain** Character. Indicator domain: quality, trace, risk, usability.

**description** Character. Detailed description.

**severity\_default** Character. Default severity level.

**source** Character. Typical source tool.

**tags** Character. Comma-separated tags.

### Source

Curated indicator definitions for the R4SUB (R for Regulatory Submission) ecosystem.

### Examples

```
data(regulatory_indicators)
table(regulatory_indicators$domain)
```

---

risk\_register\_pharma *Pharma Risk Register*

---

### Description

A Failure Mode and Effects Analysis (FMEA)-based risk register with 18 risks covering data quality, traceability, documentation, programming, and compliance categories. Includes probability, impact, and detectability scores on a 1-5 scale. Structured according to ICH (International Council for Harmonisation) Q9 quality risk management principles.

**Usage**

```
risk_register_pharma
```

**Format**

A tibble with 18 rows and 9 columns:

**risk\_id** Character. Unique risk identifier (RISK-001 to RISK-018).

**description** Character. Risk description.

**category** Character. Risk category.

**probability** Integer. Probability of occurrence (1-5).

**impact** Integer. Impact severity (1-5).

**detectability** Integer. Detectability rating (1-5).

**owner** Character. Risk owner name.

**mitigation** Character. Mitigation action (or NA).

**status** Character. Status: open, mitigated, closed, accepted.

**Source**

Synthetic risk register based on ICH (International Council for Harmonisation) Q9 quality risk management principles.

**Examples**

```
data(risk_register_pharma)
table(risk_register_pharma$category)
```

---

 sdm\_metadata

---

*SDTM Variable-Level Metadata*


---

**Description**

SDTM (Study Data Tabulation Model) variable-level metadata for DM (Demographics, 17 vars), AE (Adverse Events, 14 vars), and LB (Laboratory Results, 12 vars). Follows CDISC (Clinical Data Interchange Standards Consortium) SDTM conventions.

**Usage**

```
sdm_metadata
```

**Format**

A tibble with 43 rows and 6 columns:

**dataset** Character. SDTM domain name (DM, AE, LB).

**variable** Character. Variable name.

**label** Character. Variable label.

**type** Character. Variable type (Char or Num).

**length** Integer. Variable length.

**format** Character. SAS (Statistical Analysis System) format (or NA).

**Source**

Synthetic metadata based on CDISC SDTM (Study Data Tabulation Model) standards.

**Examples**

```
data(sdtm_metadata)
table(sdtm_metadata$dataset)
```

---

trace\_mapping

*ADaM-to-SDTM Traceability Mapping*

---

**Description**

Maps ADaM (Analysis Data Model) variables to their SDTM (Study Data Tabulation Model) source variables with derivation text and confidence scores. Includes direct copies, derived variables, and unmapped entries. Follows CDISC (Clinical Data Interchange Standards Consortium) traceability conventions.

**Usage**

```
trace_mapping
```

**Format**

A tibble with 25 rows and 6 columns:

**adam\_dataset** Character. Source ADaM dataset.

**adam\_var** Character. Source ADaM variable.

**sdtm\_domain** Character. Target SDTM domain (NA if derived).

**sdtm\_var** Character. Target SDTM variable (NA if derived).

**derivation\_text** Character. Derivation description text.

**confidence** Numeric. Mapping confidence score (0-1, NA if unmapped).

**Source**

Synthetic traceability mapping based on CDISC conventions.

**Examples**

```
data(trace_mapping)  
table(trace_mapping$adam_dataset)
```

# Index

## \* datasets

- adam\_metadata, 2
- evidence\_pharma, 3
- regulatory\_indicators, 5
- risk\_register\_pharma, 5
- sdtm\_metadata, 6
- trace\_mapping, 7

adam\_metadata, 2

dataset\_dictionary, 3

evidence\_pharma, 3

list\_datasets, 4

regulatory\_indicators, 5

risk\_register\_pharma, 5

sdtm\_metadata, 6

trace\_mapping, 7