Package: HCUPtools (via r-universe)

October 8, 2026

Type Package

Title Access and Work with HCUP Resources and Datasets

Version 1.0.2

Depends R (>= 4.1.0)

Description A comprehensive R package for accessing and working with publicly available and free resources from the Agency for Healthcare Research and Quality (AHRQ) Healthcare Cost and Utilization Project (HCUP). The package provides streamlined access to HCUP's Clinical Classifications Software Refined (CCSR) mapping files and Summary Trend Tables, enabling researchers and analysts to efficiently map ICD-10-CM diagnosis codes and ICD-10-PCS procedure codes to CCSR categories and access HCUP statistical reports. Key features include: direct download from HCUP website, multiple output formats (long/wide/default), cross-classification support, version management, citation generation, and intelligent caching. The package does not redistribute HCUP data files but facilitates direct download from the official HCUP website, ensuring users always have access to the latest versions and maintain compliance with HCUP data use policies. This package only accesses free public tools and reports; it does NOT access HCUP databases (NIS, KID, SID, NEDS, etc.) that require purchase. For more information, see <https://hcup-us.ahrq.gov/>.

License MIT + file LICENSE

URL https://github.com/vikrant31/HCUPtools, https://vikrant31.github.io/HCUPtools/

BugReports https://github.com/vikrant31/HCUPtools/issues

Encoding UTF-8

LazyData true

Roxygen list(markdown = TRUE)

RoxygenNote 7.3.3

Imports httr2, readr, dplyr, tidyr, tibble, utils, stats, rlang, xml2, readxl

Suggests testthat (>= 3.0.0), knitr, rmarkdown, data.table, pdftools

VignetteBuilder knitr

Config/pak/sysreqs libicu-dev libxml2-dev libssl-dev libx11-dev

Repository https://vikrant31.r-universe.dev

Date/Publication 2026-07-13 01:29:25 UTC

RemoteUrl https://github.com/vikrant31/hcuptools

RemoteRef HEAD

RemoteSha b85dc62d6d9ed0330038071b08f2411886dd953f

Contents

ccsr_changelog 2
ccsr_map 4
download_ccsr 6
download_trend_tables 7
get_ccsr_description 9
hcup_citation 10
list_ccsr_versions 11
list_trend_table_sheets 12
read_ccsr 12
read_trend_table 14

Index **17**

`ccsr_changelog` *Get CCSR Change Log*

Description

Retrieves and displays the change log for CCSR versions. The change log documents updates, additions, and modifications to CCSR categories across different versions.

Usage

```
ccsr_changelog(
  version = "latest",
  type = "diagnosis",
  format = "read",
  as_data_table = NULL
)
```

Arguments

`version`	Character string specifying the CCSR version. Use "latest" (default) to get the change log for the most recent version, or specify a version like "v2026.1", "v2025.1", etc.
`type`	Character string specifying the type of CCSR. Must be one of: "diagnosis" (or "dx") for ICD-10-CM diagnosis codes, or "procedure" (or "pr") for ICD-10-PCS procedure codes. Default is "diagnosis".
`format`	Character string specifying the output format. Options: "read" (default) - Downloads and reads the Excel file as a data table/tibble (requires `readxl` package) "text" - Returns change log information as text "url" - Returns the URL to the change log document "download" - Downloads and returns the change log file path "view" - Downloads and opens the change log file in the default viewer "extract" - Attempts to extract text from the file (requires `pdftools` for PDF or `readxl` for Excel)
`as_data_table`	Logical. If TRUE, returns a `data.table` instead of a tibble. Only used when `format = "read"`. If NULL (default), prompts the user interactively to choose (only in interactive sessions). In non-interactive sessions, defaults to FALSE.

Details

CCSR change logs document:

- New CCSR categories added
- Categories that were removed or merged
- Changes to category descriptions
- Updates to ICD-10 code mappings
- Version-specific notes and improvements

Change logs are typically available as PDF or text documents on the HCUP website. This function attempts to locate and retrieve them.

Value

Depending on `format`:

- "read" (default): A tibble or data.table containing the change log data (if Excel file)
- "text": Character string with change log information
- "url": Character string with URL to change log
- "download": Character string with path to downloaded file
- "view": Opens the file and returns the file path (invisibly)
- "extract": Character string with extracted text from file

Examples

```
# Get latest change log URL
changelog_url <- ccsr_changelog(format = "url")

# Get change log information
changelog_info <- ccsr_changelog(version = "v2026.1", format = "text")

# Download change log file
changelog_file <- ccsr_changelog(version = "v2025.1", format = "download")

# View change log in default PDF viewer
ccsr_changelog(version = "v2026.1", format = "view")

# Extract text from change log PDF (requires pdftools package)
changelog_text <- ccsr_changelog(version = "v2026.1", format = "extract")
cat(changelog_text)
```

---

ccsr_map *Map ICD-10 Codes to CCSR Categories*

Description

Maps ICD-10-CM diagnosis codes or ICD-10-PCS procedure codes to their corresponding CCSR categories using a downloaded CCSR mapping file.

Usage

```
ccsr_map(
  data,
  code_col,
  map_df,
  type = NULL,
  default_only = FALSE,
  output_format = "long",
  keep_all = TRUE
)
```

Arguments

data	A data frame or tibble containing ICD-10 codes to be mapped.
code_col	Character string specifying the name of the column in data that contains the ICD-10 codes.
map_df	A tibble containing the CCSR mapping data, typically obtained from download_ccsr().

type	Character string specifying the type of mapping. Must be one of: "diagnosis" (or "dx") for ICD-10-CM codes, or "procedure" (or "pr") for ICD-10-PCS codes. If NULL (default), the function will attempt to infer the type from the mapping data frame.
default_only	Logical. For diagnosis codes only, if TRUE, returns only the default CCSR category (recommended for principal diagnosis analysis). If FALSE (default), returns all assigned CCSR categories including cross-classifications.
output_format	Character string specifying the output format. Must be one of: "long" (default) or "wide". "long" format duplicates records for each assigned CCSR category. "wide" format creates multiple columns (CCSR_1, CCSR_2, etc.) for multiple categories.
keep_all	Logical. If TRUE (default), returns all original columns from data plus the CCSR mapping columns. If FALSE, returns only the ICD-10 code column and CCSR mapping columns.

Details

CCSR allows for cross-classification, meaning a single ICD-10 code can map to multiple CCSR categories. The "long" format is recommended for analyses where you want to count all assigned CCSR categories, while "wide" format may be more convenient for patient-level analyses.

For diagnosis codes, CCSR also assigns a "default" category that is recommended for principal diagnosis analysis. Use default_only = TRUE to extract only this default category.

Value

A tibble with the original data plus CCSR mapping columns. The structure depends on output_format:

- For "long" format: Each row represents one ICD-10 code and one CCSR category assignment (rows are duplicated for multiple categories).
- For "wide" format: Each row represents one ICD-10 code with multiple CCSR category columns (CCSR_1, CCSR_2, etc.).

Examples

```
# Download mapping file
dx_map <- download_ccsr("diagnosis")

# Create sample data
sample_data <- tibble::tibble(
  patient_id = 1:3,
  icd10_code = c("E11.9", "I10", "M79.3")
)

# Map codes (long format - default)
mapped_long <- ccsr_map(
  data = sample_data,
  code_col = "icd10_code",
  map_df = dx_map
)
```

```
# Map codes (wide format)
mapped_wide <- ccsr_map(
  data = sample_data,
  code_col = "icd10_code",
  map_df = dx_map,
  output_format = "wide"
)

# Map codes (default category only)
mapped_default <- ccsr_map(
  data = sample_data,
  code_col = "icd10_code",
  map_df = dx_map,
  default_only = TRUE
)
```

`download_ccsr` *Download CCSR Mapping Files from HCUP*

Description

Downloads and loads Clinical Classifications Software Refined (CCSR) mapping files directly from the Agency for Healthcare Research and Quality (AHRQ) Healthcare Cost and Utilization Project (HCUP) website.

Usage

```
download_ccsr(
  type = "diagnosis",
  version = "latest",
  cache = TRUE,
  clean_names = TRUE
)
```

Arguments

`type`	Character string specifying the type of CCSR file to download. Must be one of: "diagnosis" (or "dx") for ICD-10-CM diagnosis codes, or "procedure" (or "pr") for ICD-10-PCS procedure codes. Default is "diagnosis".
`version`	Character string specifying the CCSR version to download. Use "latest" to download the most recent version, or specify a version like "v2026.1", "v2025.1", etc. Default is "latest".
`cache`	Logical. If TRUE (default), the downloaded file is cached in a temporary directory to avoid re-downloading on subsequent calls.
`clean_names`	Logical. If TRUE (default), column names are cleaned to follow R naming conventions (snake_case).

Details

This function downloads CCSR mapping files directly from the HCUP website. The package does not redistribute these files but facilitates access to the official AHRQ data sources.

The function handles:

- Automatic URL construction based on type and version
- ZIP file download and extraction
- Proper encoding of special characters
- Preservation of leading zeros in ICD-10 codes
- Conversion to tidy tibble format

Value

A tibble containing the CCSR mapping data with the following columns:

- For diagnosis files: ICD-10-CM code, CCSR category, default CCSR category, and clinical descriptions
- For procedure files: ICD-10-PCS code, CCSR category, and descriptions

Examples

```
# Download latest diagnosis CCSR mapping
dx_map <- download_ccsr("diagnosis")

# Download specific version of procedure CCSR mapping
pr_map <- download_ccsr("procedure", version = "v2025.1")

# Download without caching
dx_map <- download_ccsr("diagnosis", cache = FALSE)
```

---

download_trend_tables *Download HCUP Summary Trend Tables*

Description

Downloads HCUP Summary Trend Tables from the HCUP website. These tables provide information on hospital utilization derived from HCUP databases, including trends in inpatient and emergency department utilization.

Usage

```
download_trend_tables(table_id = NULL, dest_dir = NULL, cache = TRUE)
```

Arguments

table_id	Character string or numeric specifying which table to download. Can be: • A table number (e.g., "1", "2a", "6b", "11c") for specific tables • "all" to download all available tables as a ZIP file • NULL (default) to show an interactive menu for selecting a table (in interactive sessions), or return a list of available tables (in non-interactive sessions)
dest_dir	Character string specifying the destination directory for the downloaded file(s). If NULL (default), files are saved to a temporary directory.
cache	Logical. If TRUE (default), downloaded files are cached to avoid re-downloading on subsequent calls.

Details

The HCUP Summary Trend Tables include information on:

- Overview of trends in inpatient and emergency department utilization
- All inpatient encounter types
- Inpatient encounter types (normal newborns, deliveries, elective/non-elective stays)
- Inpatient service lines (maternal/neonatal, mental health, injuries, surgeries, etc.)
- ED treat-and-release visits

Each table is available as an Excel file with state-specific, region-specific, and national statistics.

The function automatically discovers available tables by scraping the HCUP website, so it will automatically adapt to new tables or version changes.

For more information, see: https://hcup-us.ahrq.gov/reports/trendtables/summarytrendtables.jsp

Value

If table_id is NULL and session is non-interactive, returns a data frame listing available tables. Otherwise, returns the path(s) to the downloaded file(s).

Examples

```
# List available tables
available_tables <- download_trend_tables()
print(available_tables)

# Download a specific table
table_path <- download_trend_tables("2a")

# Bulk ZIP (\code{table_id = "all"}) is only available when HCUP publishes a
# combined file; otherwise use individual IDs from the listing above.
```

`get_ccsr_description` *Get CCSR Category Descriptions*

Description

Retrieves the full clinical description for one or more CCSR category codes. This function helps users interpret CCSR codes by providing their meaningful clinical descriptions.

Usage

```
get_ccsr_description(ccsr_codes, map_df = NULL, type = NULL)
```

Arguments

`ccsr_codes`	Character vector of CCSR category codes (e.g., "ADM010", "NEP003", "CIR019").
`map_df`	Optional. A tibble containing CCSR mapping data with descriptions. If provided, descriptions are extracted from this data frame. If NULL (default), the function will attempt to download the latest mapping file to extract descriptions.
`type`	Character string specifying the type of CCSR codes. Must be one of: "diagnosis" (or "dx") or "procedure" (or "pr"). If NULL (default), the function will attempt to infer the type from the codes or mapping data.

Details

CCSR category codes follow specific naming conventions:

- Diagnosis codes: Typically start with letters (e.g., "ADM010", "NEP003")
- Procedure codes: Typically start with letters (e.g., "PRC001", "PRC002")

If a description is not found for a code, it will be marked as NA in the result.

Value

A tibble with columns:

- `ccsr_code`: The CCSR category code
- `description`: The full clinical description
- Additional metadata columns if available in the mapping data

Examples

```
# Get descriptions using downloaded mapping data
dx_map <- download_ccsr("diagnosis")
get_ccsr_description(c("ADM010", "NEP003", "CIR019"), map_df = dx_map)

# Get descriptions without pre-downloaded data (will download automatically)
get_ccsr_description(c("ADM010", "NEP003"), type = "diagnosis")
```

`hcup_citation` *Generate Citation for HCUP Resources*

Description

Provides recommended citations for HCUP resources including Clinical Classifications Software Refined (CCSR) data and Summary Trend Tables from the Agency for Healthcare Research and Quality (AHRQ) Healthcare Cost and Utilization Project (HCUP).

Usage

```
hcup_citation(format = "text", version = "latest", resource = "ccsr")
```

Arguments

`format`	Character string specifying the citation format. Must be one of: "text" (default), "bibtex", or "r" (for R citation object).
`version`	Character string specifying the CCSR version to cite. If "latest" (default), the function will attempt to fetch the latest version from the HCUP website. Otherwise, specify a version like "v2026.1".
`resource`	Character string specifying which HCUP resource to cite. Options: "ccsr" (default) for CCSR data, or "trend_tables" for Summary Trend Tables.

Details

This function generates citations for HCUP resources following AHRQ/HCUP guidelines. The citation includes the appropriate version number and access date. For CCSR data, the version is automatically detected if not specified. For Summary Trend Tables, the citation references the general HCUP Summary Trend Tables resource.

Value

If `format` is "text", returns a character string with the citation. If `format` is "bibtex", returns a character string with BibTeX format. If `format` is "r", returns an R citation object.

Examples

```
# Text citation for CCSR
hcup_citation()

# BibTeX format for CCSR
hcup_citation(format = "bibtex")

# Citation for Summary Trend Tables
hcup_citation(resource = "trend_tables")

# R citation object
hcup_citation(format = "r")
```

`list_ccsr_versions` *List Available CCSR Versions*

Description

Returns a list of available CCSR versions for download by scraping the HCUP website. This function helps users identify which versions are available for diagnosis and procedure mapping files.

Usage

```
list_ccsr_versions(type = "all")
```

Arguments

`type`	Character string specifying the type of CCSR file. Must be one of: "diagnosis" (or "dx"), "procedure" (or "pr"), or "all" (default) to list versions for both types.

Details

This function fetches available CCSR versions from the HCUP website. Results are cached for 24 hours to minimize website requests. If the website cannot be accessed, the function will return an error.

Value

A data frame (tibble) with columns:

- `type`: The CCSR type ("diagnosis" or "procedure")
- `version`: The version identifier (e.g., "v2026.1")

Examples

```
# List all available versions
list_ccsr_versions()

# List only diagnosis versions
list_ccsr_versions("diagnosis")

# List only procedure versions
list_ccsr_versions("procedure")
```

list_trend_table_sheets *List Available Sheets in Trend Table*

Description

Lists all available sheets in a HCUP Summary Trend Table Excel file.

Usage

```
list_trend_table_sheets(file_path)
```

Arguments

file_path	Character string, path to a trend table Excel file (.xlsx).

Value

A character vector of sheet names.

Examples

```
# Requires network: download first, then list sheets from that file path
path_xlsx <- download_trend_tables("2a")
list_trend_table_sheets(path_xlsx)
```

read_ccsr *Read CCSR Mapping Files from Disk*

Description

Reads previously downloaded CCSR mapping files from disk. If no file path is provided, automatically finds and reads cached files from `download_ccsr()`.

Usage

```
read_ccsr(
  file_path = NULL,
  type = NULL,
  version = "latest",
  clean_names = TRUE,
  as_data_table = NULL,
  name = NULL
)
```

Arguments

file_path	Optional character string, path to a CCSR mapping file. Can be: • A ZIP file path (will be extracted and read) • A CSV/Excel file path (will be read directly) • A directory path containing extracted CCSR files If NULL (default), automatically searches the cache directory for files downloaded via download_ccsr().
type	Character string specifying the type of CCSR file. Must be one of: "diagnosis" (or "dx") for ICD-10-CM diagnosis codes, or "procedure" (or "pr") for ICD-10-PCS procedure codes. If NULL and file_path is NULL, defaults to "diagnosis". If file_path is provided, the function will attempt to infer the type from the file name or contents.
version	Character string specifying the CCSR version to read from cache. Use "latest" (default) to read the most recent version, or specify a version like "v2026.1", "v2025.1", etc. Only used when file_path is NULL.
clean_names	Logical. If TRUE (default), column names are cleaned to follow R naming conventions (snake_case).
as_data_table	Logical or NULL. If TRUE and the data.table package is available, returns a data.table object instead of a tibble. If FALSE, returns a tibble. If NULL (default), prompts the user interactively to choose (only in interactive sessions). In non-interactive sessions, defaults to FALSE. Note: tibbles are already data frames and work with all standard R data frame operations.
name	Optional character string, suggested variable name for the returned data. This is only used for display/messaging purposes and does not automatically assign the data to a variable. You must still assign the result: my_data <- read_ccsr(). If NULL (default), a name is suggested based on the file type and version.

Details

This function can read CCSR files in several formats:

- ZIP files downloaded from HCUP (will extract and read the CSV/Excel file)
- CSV files (extracted from ZIP or saved separately)
- Excel files (if readxl package is available)
- Directories containing extracted files
- Cached files from download_ccsr() (automatic if file_path is NULL)

The function automatically detects the file format and handles encoding issues, preserving leading zeros in ICD-10 codes.

When file_path is NULL, the function automatically searches the cache directory (tempdir()/HCUPtools_cache/) for files matching the specified type and version. This makes it easy to read previously downloaded files without needing to know the exact file path.

Value

A tibble (or data.table if as_data_table = TRUE) containing the CCSR mapping data. Tibbles are data frames and can be used with all standard R data frame operations, including dplyr, data.table, and base R functions.

Note

To use the data, assign it to a variable: `my_data <- read_ccsr()`. The `name` parameter is only for display purposes and does not automatically assign the data.

Examples

```
# Populate cache, then read (requires network)
invisible(download_ccsr("diagnosis"))
dx_map <- read_ccsr(as_data_table = FALSE)
invisible(download_ccsr("procedure"))
pr_map <- read_ccsr(type = "procedure", as_data_table = FALSE)

# From a local file on your machine (uncomment and set the path):
# dx_map <- read_ccsr("/path/to/DXCCSR-v2026-1.zip", as_data_table = FALSE)
# dx_map <- read_ccsr("/path/to/DXCCSR_v2026_1.csv", as_data_table = FALSE)
# dx_map <- read_ccsr("/path/to/extracted_ccsr_files/", as_data_table = FALSE)

head(dx_map)
nrow(dx_map)
```

---

read_trend_table *Read HCUP Summary Trend Table from Disk*

Description

Reads a previously downloaded HCUP Summary Trend Table Excel file from disk. If no file path is provided, automatically finds and reads cached files from `download_trend_tables()`, with an interactive menu to select from available tables.

Usage

```
read_trend_table(
  file_path = NULL,
  table_id = NULL,
  sheet = NULL,
  clean_names = TRUE,
  as_data_table = NULL,
  name = NULL
)
```

Arguments

`file_path`	Optional character string, path to a trend table Excel file (.xlsx). If NULL (default), automatically searches the cache directory for files downloaded via `download_trend_tables()` and shows an interactive menu.

table_id	Optional character string, table ID (e.g., "1", "2a", "2b") to read from cache. Only used when file_path is NULL. If both are NULL, shows interactive menu.
sheet	Character string or integer specifying which sheet to read. If NULL (default), shows an interactive menu to select a sheet (in interactive sessions), or automatically selects the "National" sheet (or first data sheet) in non-interactive sessions. Common sheet names include "National", "Regional", "State", etc.
clean_names	Logical. If TRUE (default), column names are cleaned to follow R naming conventions (snake_case).
as_data_table	Logical or NULL. If TRUE and the data.table package is available, returns a data.table object instead of a tibble. If FALSE, returns a tibble. If NULL (default), prompts the user interactively to choose (only in interactive sessions). In non-interactive sessions, defaults to FALSE. Note: tibbles are already data frames and work with all standard R data frame operations.
name	Optional character string, suggested variable name for the returned data. This is only used for display/messaging purposes and does not automatically assign the data to a variable. You must still assign the result: my_data <- read_trend_table(). If NULL (default), a name is suggested based on the table ID and sheet.

Details

HCUP Summary Trend Tables are Excel files with multiple sheets containing data at different geographic levels (National, Regional, State). Use the sheet parameter to specify which sheet to read, or call the function multiple times with different sheets.

When file_path is NULL, the function automatically searches the cache directory (tempdir()) for files matching the pattern HCUP_SummaryTrendTables_*.xlsx. If multiple files are found, an interactive menu is displayed for selection.

To see available sheets, use list_trend_table_sheets().

Value

A tibble (or data.table if as_data_table = TRUE) containing the trend table data. Tibbles are data frames and can be used with all standard R data frame operations, including dplyr, data.table, and base R functions.

Note

To use the data, assign it to a variable: my_data <- read_trend_table(). The name parameter is only for display purposes and does not automatically assign the data.

Examples

```
# Requires network: download a table, list sheets, read data (same file path)
path_xlsx <- download_trend_tables("2a")
list_trend_table_sheets(path_xlsx)
national_data <- read_trend_table(file_path = path_xlsx, as_data_table = FALSE)
head(national_data)
```

```
# After a download, you can also read from cache by table ID
table_2a <- read_trend_table(table_id = "2a", as_data_table = FALSE)
head(table_2a)

# With a file already on disk, pass its path to `read_trend_table(file_path = ...)`.
```

Index

ccsr_changelog, 2
ccsr_map, 4

download_ccsr, 6
download_trend_tables, 7

get_ccsr_description, 9

hcup_citation, 10

list_ccsr_versions, 11
list_trend_table_sheets, 12

read_ccsr, 12
read_trend_table, 14