

# Package ‘cthist’

July 17, 2024

**Title** Clinical Trial Registry History

**Version** 2.1.11

**Description** Retrieves historical versions of clinical trial registry entries from <https://ClinicalTrials.gov>. Package functionality and implementation for v 1.0.0 is documented in Carlisle (2022) [DOI:10.1371/journal.pone.0270909](https://doi.org/10.1371/journal.pone.0270909).

**License** AGPL (>= 3)

**URL** <https://github.com/bgcarlisle/cthist>

**BugReports** <https://github.com/bgcarlisle/cthist/issues>

**Imports** assertthat, dplyr, httr, jsonlite, lubridate, magrittr, readr, rlang, stringr, tibble, purrr, tidyr

**Suggests** testthat (>= 3.0.0)

**Config/testthat/edition** 3

**Encoding** UTF-8

**RoxygenNote** 7.3.2

**NeedsCompilation** no

**Author** Benjamin Gregory Carlisle [aut, cre]  
(<https://orcid.org/0000-0001-8975-0649>)

**Maintainer** Benjamin Gregory Carlisle <murph@bgcarlisle.com>

**Repository** CRAN

**Date/Publication** 2024-07-17 16:10:02 UTC

## Contents

clinicaltrials_gov_dates . . . . .	2
clinicaltrials_gov_download . . . . .	2
clinicaltrials_gov_version . . . . .	4
extract_publications . . . . .	5
overall_status_lengths . . . . .	6
<b>Index</b>	<b>7</b>

clinicaltrials\_gov\_dates

*Download a table of dates on which a ClinicalTrials.gov registry entry was updated*

---

### **Description**

Download a table of dates on which a ClinicalTrials.gov registry entry was updated

### **Usage**

```
clinicaltrials_gov_dates(nctids, status_change_only = FALSE, quiet = TRUE)
```

### **Arguments**

**nctids** A list of well-formed NCT numbers, e.g. c("NCT00942747", "NCT03281616"). (A capitalized "NCT" followed by eight numerals with no spaces or hyphens.)

**status\_change\_only** If TRUE, returns only the dates marked with a Recruitment Status change, default FALSE.

**quiet** A boolean TRUE or FALSE. If TRUE, no messages will be printed during download. TRUE by default, messages printed for every registry entry downloaded showing progress.

### **Value**

A table with three columns: the version number (starting from 0), the ISO-8601 formatted date on which there were clinical trial history version updates, and the trial's overall status on that date.

### **Examples**

```
versions <- clinicaltrials_gov_dates("NCT00942747")
```

---

clinicaltrials\_gov\_download

*Mass-download registry entry historical versions from ClinicalTrials.gov*

---

## Description

This function will download all ClinicalTrials.gov registry records for the NCT numbers specified. Rather than transcribing NCT numbers by hand, it is recommended that you conduct a search for trials of interest using the ClinicalTrials.gov web front-end and download the result as a comma-separated value (CSV) file. The CSV can be read in to memory as a data frame and the NCT Number column can be passed directly to the function as the `nctids` argument.

## Usage

```
clinicaltrials_gov_download(  
  nctids,  
  output_filename = NA,  
  quiet = FALSE,  
  earliest = FALSE,  
  latest = FALSE  
)
```

## Arguments

<code>nctids</code>	A list of well-formed NCT numbers, e.g. <code>c("NCT00942747", "NCT03281616")</code> .
<code>output_filename</code>	A character string for a filename into which the data frame will be written as a CSV, e.g. "historical_versions.csv". If no output filename is provided, the data frame of downloaded historical versions will be returned by the function as a data frame.
<code>quiet</code>	A boolean TRUE or FALSE. If TRUE, no messages will be printed during download. FALSE by default, messages printed for every version downloaded showing progress.
<code>earliest</code>	A boolean TRUE or FALSE. If TRUE, only the earliest version of the registry entry will be downloaded, if FALSE, all versions will be downloaded. FALSE by default. Can be combined with latest.
<code>latest</code>	A boolean TRUE or FALSE. If TRUE, only the latest version of the registry entry will be downloaded, if FALSE, all versions will be downloaded. FALSE by default. Can be combined with earliest.

## Value

If an output filename is specified, on successful completion, this function returns TRUE and otherwise returns FALSE. If an output filename is not specified, on successful completion, this function returns a data frame containing the historical versions of the clinical trial that have been retrieved, and in case of error returns FALSE. After unsuccessful completion with an output filename specified, if the function is called again with the same NCT numbers and output filename, the function will check the output file for errors or incompletely downloaded registry entries, remove them and try to download the historical versions that are still needed, while preserving the ones that have already been downloaded correctly.

**Examples**

```
filename <- tempfile()
clinicaltrials_gov_download(c("NCT00942747",
                             "NCT03281616"), filename)
```

```
hv <- clinicaltrials_gov_download("NCT00942747")
```

---

clinicaltrials\_gov\_version

*Download a registry entry version from ClinicalTrials.gov*

---

**Description**

Download a registry entry version from ClinicalTrials.gov

**Usage**

```
clinicaltrials_gov_version(nctid, versionno = 0)
```

**Arguments**

nctid	A character string including a well-formed ClinicalTrials.gov NCT Number, e.g. "NCT00942747". (A capitalized "NCT" followed by eight numerals with no spaces or hyphens.)
versionno	An integer version number, e.g. 3, where 0 is the earliest version of the trial in question, 1 is the next most recent, etc. (Please note that this differs from the convention used in cthist v. <= 1.4.2, in which 1 is the earliest version of the trial in question.) If no version number is specified, the first version will be downloaded. If -1 (negative one) is specified, the latest version will be downloaded.

**Value**

A list containing the overall status, enrolment, start date, start date precision (month or day) primary completion date, primary completion date precision (month or day), primary completion date type, minimum age, maximum age, sex, accepts healthy volunteers, inclusion/exclusion criteria, outcome measures, overall contacts, central contacts, responsible party, lead sponsor, collaborators, locations, reason why the trial stopped (if provided), whether results are posted, references data, organization identifiers and other secondary trial identifiers.

**Examples**

```
version <- clinicaltrials_gov_version("NCT00942747", 1)
```

---

extract_publications	<i>Takes a data frame of the type provided by <code>clinicaltrials_gov_download()</code> and returns a new data frame containing one row per publication of the publication type specified indexed on <code>ClinicalTrials.gov</code> for every version of the clinical trial record provided.</i>
----------------------	--

---

### Description

This function does not connect to `ClinicalTrials.gov`, and only interprets data that has already been downloaded by expanding the nested JSON-encoded data in the `references` column provided by `clinicaltrial_gov_version`.

### Usage

```
extract_publications(df, types = c("RESULT", "BACKGROUND", "DERIVED"))
```

### Arguments

<code>df</code>	A data frame containing at least the following columns: <code>nctid</code> , <code>version_number</code> , <code>total_versions</code> , <code>version_date</code> , and <code>references</code> . The <code>references</code> column should contain a nested JSON-encoded table with three columns: <code>pmid</code> , <code>type</code> and <code>citation</code> . This data frame can be generated by the use of <code>clinicaltrials_gov_download</code> .
<code>types</code>	A list of types to be returned or a character string if only one type specified, e.g. "RESULT" or <code>c("RESULT", "BACKGROUND")</code> . Allowed types: "RESULT", "BACKGROUND", "DERIVED".

### Value

A data frame with all the original columns, as well as an additional three columns: `pmid`, `type` and `citation`. The new data frame will have one row per publication.

### Examples

```
hv <- clinicaltrials_gov_download("NCT00942747", latest=TRUE)
extract_publications(hv)
```

---

overall\_status\_lengths

*Interpret downloaded version histories to determine how long in days a trial had any given overall status*

---

### Description

This function takes a data frame of the type produced by `clinicaltrials_gov_download()` or `clinicaltrials_gov_dates()` and interprets it to determine, for each clinical trial registry entry, how many days were spent in each overall status (e.g. "RECRUITING", "ACTIVE, NOT RECRUITING", etc.); upper and lower date bounds can also be applied, to allow for returning only those dates that fall within a time range of interest.

### Usage

```
overall_status_lengths(  
  historical_versions,  
  start_date = NA,  
  end_date = NA,  
  carry_forward_last_status = TRUE  
)
```

### Arguments

`historical_versions`  
A data frame of the type produced by `clinicaltrials_gov_download()` or `clinicaltrials_gov_dates()`. Must include a row for every historical version, with the `nctid` column specifying the clinical trial registry entry, the `overall_status` column indicating the status of the trial, and the `version_date` column indicating the date on which the registry entry was updated. Other columns optional.

`start_date`  
A date or character string in YYYY-MM-DD format specifying a date. If specified, only the length of time that is after the given start date will be counted.

`end_date`  
A date or character string in YYYY-MM-DD format specifying a date. If specified, only the length of time that is before the given end date will be counted.

`carry_forward_last_status`  
Boolean TRUE or FALSE.

### Value

A data frame with two columns: `nctid`, which contains all the distinct NCT numbers from the `historical_versions` data frame provided, and `days`, which contains the number of

# Index

clinicaltrials\_gov\_dates, [2](#)  
clinicaltrials\_gov\_download, [2](#)  
clinicaltrials\_gov\_version, [4](#)  
  
extract\_publications, [5](#)  
  
overall\_status\_lengths, [6](#)