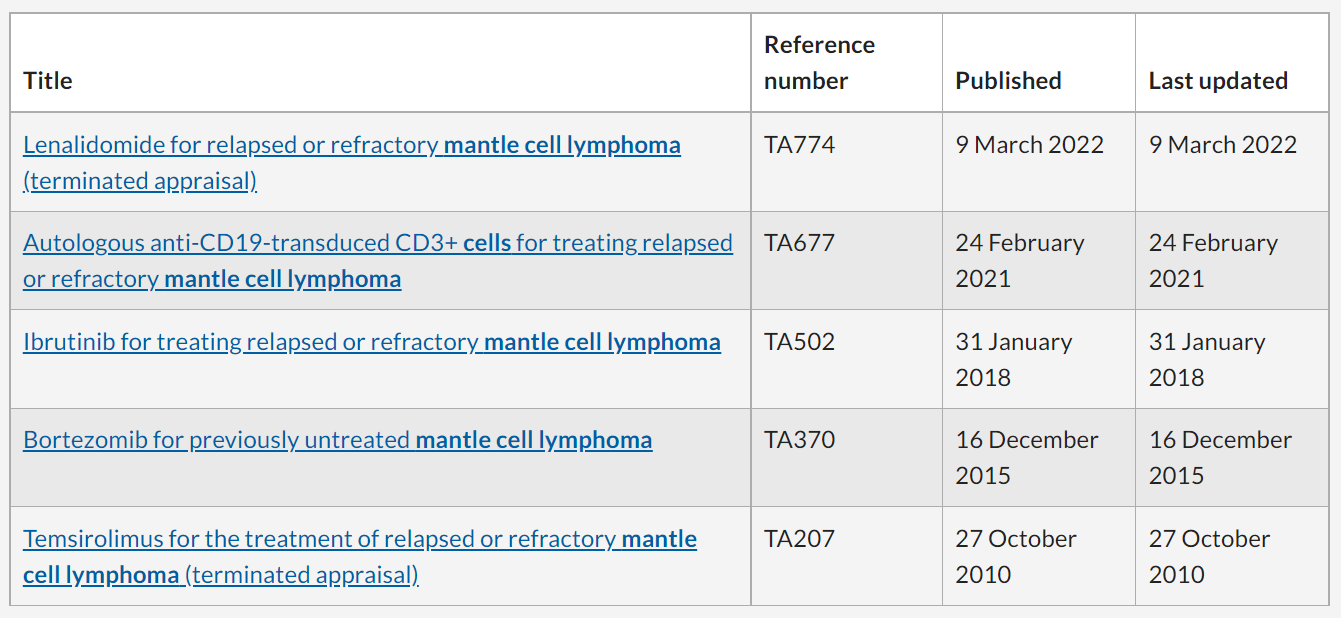
**READ ME coding notes**

* Extract references from NICE website (they can be copied-and-pasted as a table from [here](https://www.nice.org.uk/guidance/published), not sure if there's a way of doing it from a search)
* Import into R studio as table
* Convert to text string
* Export as text file

Here's how references display on the NICE website, using Tom's link above



The R Ladies of Sydney suggest finding a package that can do the task you are working on rather than creating code yourself, if possible. This package is on the CRAN list:

RIO (a Swiss army knife for data)

[Import, Export, and Convert Data Files (r-project.org)](https://cran.r-project.org/web/packages/rio/vignettes/rio.html)

Is there a better way to get references out of the website than direct search?

Importing a CSV file into Rstudio - I found the following article/blog, one section of which includes:

"[Importing data from text and CSV files](https://support.rstudio.com/hc/en-us/articles/218611977-Importing-Data-with-the-RStudio-IDE#importing-csv)"

**Week 1 notes**

Downloadable files in Excel from NICE website - useful to look at this in terms of upskilling exercise - but may not be the ideal way of doing it given the amount of downloading, internal searching required. Using APIs may be the better way, potentially setting up a web app for people to use to get their formatted guidelines

Potentially a route through the Evidence search replacement

First we need to manipulate the data - can we work with website provider to make it more fit for our purposes.

Need to map out the transformations from published format

We need to import (Excel or CSV),

Title goes first

Hyperlink needs to be carried over - CSV may not carry the hyperlink

* <https://community.rstudio.com/t/how-to-preserve-hyperlinks-in-r-htmltable-while-writing-to-a-csv-excel-file/37728/2>
* DN to check on CSV output and hyperlink display

Conditional rule - if it starts with QS its a quality standard - variables for clinical guidelines due to change of numbering system

Date of publication

Update dates - may say publication date - if something matches then do not add an update date

**Week 2 notes**

Tom's tiday annotated script from Week 2

# access tidyverse library

library(tidyverse)

# import CSV file

published\_guidelines <- read\_csv("L:/IR/\_IS/Learning & Dev/Coding group/Current projects/Rough work/Tom/220228 GAQS pub.csv")

# export as .txt (just to see if we can)

write.table(published\_guidelines,file = "published\_guidelines.txt",row.names = FALSE)

# create new dataframe, a version of the old one with changed column order. Note that column 1 header is title of the original CSV and other columns only have numbers

published\_guidelines\_reordered <- select(published\_guidelines, "Published guidance, NICE advice and quality standards", 3, 4, 2)

# change published\_guidelines\_reordered to new object pgr\_ng. pgr\_ng has the letters NG in column 4 replaced with the text "NICE guideline NG"

pgr\_ng <-

published\_guidelines\_reordered %>%

mutate(across(4, str\_replace, "^NG", "NICE guideline NG"))

pgr\_ng

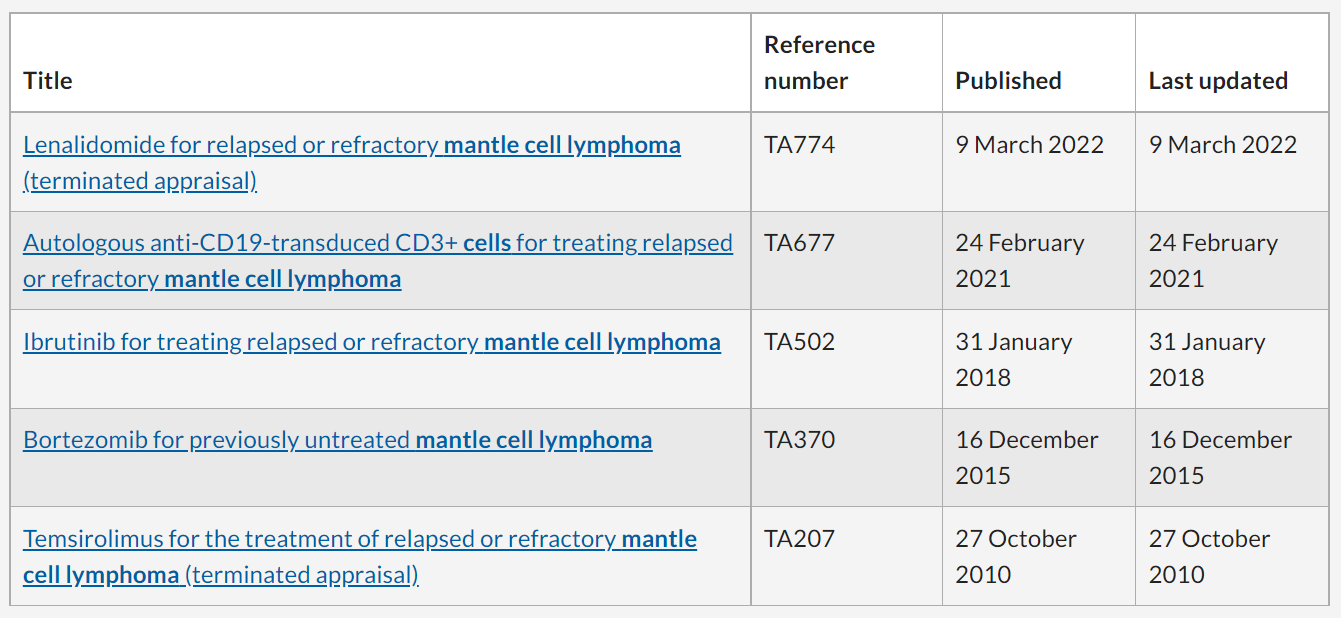
02 March 2023

10:44

Fields to extract

* Extract references from NICE website (they can be copied-and-pasted as a table from [here](https://www.nice.org.uk/guidance/published), not sure if there's a way of doing it from a search)
* Import into R studio as table
* Convert to text string
* Export as text file

Here's how references display on the NICE website, using Tom's link above



The R Ladies of Sydney suggest finding a package that can do the task you are working on rather than creating code yourself, if possible. This package is on the CRAN list:

RIO (a Swiss army knife for data)

[Import, Export, and Convert Data Files (r-project.org)](https://cran.r-project.org/web/packages/rio/vignettes/rio.html)

[NICE naming conventions for guidance](file:///L:/IR/_IS/Methods%20Guides/Format%20for%20recording%20NICE%20guidance%202020.docx)

Is there a better way to get references out of the website than direct search?

Importing a CSV file into Rstudio - I found the following article/blog, one section of which includes:

"[Importing data from text and CSV files](https://support.rstudio.com/hc/en-us/articles/218611977-Importing-Data-with-the-RStudio-IDE#importing-csv)"

**Week 1 notes**

Downloadable files in Excel from NICE website - useful to look at this in terms of upskilling exercise - but may not be the ideal way of doing it given the amount of downloading, internal searching required. Using APIs may be the better way, potentially setting up a web app for people to use to get their formatted guidelines

Potentially a route through the Evidence search replacement

First we need to manipulate the data - can we work with website provider to make it more fit for our purposes.

Need to map out the transformations from published format

We need to import (Excel or CSV),

Title goes first

Hyperlink needs to be carried over - CSV may not carry the hyperlink

* <https://community.rstudio.com/t/how-to-preserve-hyperlinks-in-r-htmltable-while-writing-to-a-csv-excel-file/37728/2>
* DN to check on CSV output and hyperlink display

Conditional rule - if it starts with QS its a quality standard - variables for clinical guidelines due to change of numbering system

Date of publication

Update dates - may say publication date - if something matches then do not add an update date

**Week 2 notes**

Tom's tiday annotated script from Week 2

# access tidyverse library

library(tidyverse)

# import CSV file

published\_guidelines <- read\_csv("L:/IR/\_IS/Learning & Dev/Coding group/Current projects/Rough work/Tom/220228 GAQS pub.csv")

# export as .txt (just to see if we can)

write.table(published\_guidelines,file = "published\_guidelines.txt",row.names = FALSE)

# create new dataframe, a version of the old one with changed column order. Note that column 1 header is title of the original CSV and other columns only have numbers

published\_guidelines\_reordered <- select(published\_guidelines, "Published guidance, NICE advice and quality standards", 3, 4, 2)

# change published\_guidelines\_reordered to new object pgr\_ng. pgr\_ng has the letters NG in column 4 replaced with the text "NICE guideline NG"

pgr\_ng <-

published\_guidelines\_reordered %>%

mutate(across(4, str\_replace, "^NG", "NICE guideline NG"))

pgr\_ng

**Week 3**

|  |  |  |  |
| --- | --- | --- | --- |
| **Fields** | **Example** | **Tag (html)** | **Notes** |
| Guideline title | Selpercatinib for previously treated RET fusion-positive advanced non-small-cell lung cancer |  |  |
| Date | The date on the webpage is: 12 January 2022. However, when we create a bibliography we usually just include the date (in brackets) e.g. (2022)    The format document for NICE guidance also has a date and update date in brackets e.g. (2015 updated 2019)    For guidelines in development there is also a "Publication expected [Month Year] [or] Publication date to be confirmed" field |  |  |
| Type of guidance | Technology appraisal guidance |  |  |
| Guideline number | The number on the website is displayed as [TA760] but the bibliography format is: 760 |  |  |
| Hyperlink | In a bibliography the guideline title contains the hyperlink |  |  |
| Other | For TAs that have been suspended we usually include text that explains why.  Status field is where we find "suspended"  Timeline update field is where we find the rationale |  |  |
|  | For the following would we do this as part of R or would it just be easier to format them once they are in Word? |  |  |
| Spacing | Do we need to insert a line space between references in the bibliography? |  |  |
| Font | Arial 10pt? |  |  |
|  |  |  |  |