

# Chapter 9. Pragmatic clinical trials in the elderly

Data cleaning and conditions studied

Javier Mancilla Galindo

2024-05-05

## Table of contents

<b>Packages and session information</b>	<b>2</b>
<b>Initial assessment of dataset</b>	<b>3</b>
<b>Type intervention</b>	<b>4</b>
<b>Conditions</b>	<b>5</b>
Wordcloud . . . . .	5
Chronic vs acute conditions . . . . .	6
Top 10 terms . . . . .	6
<b>Package References</b>	<b>7</b>

## Packages and session information

R version 4.4.0 (2024-04-24 ucrt)  
Platform: x86\_64-w64-mingw32/x64  
Running under: Windows 11 x64 (build 22631)

Matrix products: default

locale:

[1] LC\_COLLATE=Spanish\_Mexico.utf8 LC\_CTYPE=Spanish\_Mexico.utf8  
[3] LC\_MONETARY=Spanish\_Mexico.utf8 LC\_NUMERIC=C  
[5] LC\_TIME=Spanish\_Mexico.utf8

time zone: Europe/Berlin

tzcode source: internal

attached base packages:

[1] stats graphics grDevices utils datasets methods base

other attached packages:

[1] report_0.5.8	gt_0.10.1	webshot2_0.1.1	htmlwidgets_1.6.4
[5] wordcloud2_0.2.1	tm_0.7-13	NLP_0.2-1	lubridate_1.9.3
[9] forcats_1.0.0	stringr_1.5.1	dplyr_1.1.4	purrr_1.0.2
[13] readr_2.1.5	tidyr_1.3.1	tibble_3.2.1	ggplot2_3.5.1
[17] tidyverse_2.0.0	pacman_0.5.1		

## Initial assessment of dataset

Total of studies retrieved:

n
1605

Some of these studies are observational:

Frequencies:

INTERVENTIONAL	OBSERVATIONAL
1532	73

Percentage:

INTERVENTIONAL	OBSERVATIONAL
95.5	4.5

Therefore, I will exclude observational studies. We now have a total of:

n
1532

Some of the studies declare having included older adults, but the condition under study is related to pregnancy. The count of studies after exclusion of pregnancy-related studies is:

n
1488

## Type intervention

I will create a variable for the type of intervention and save the clean dataset.

behavioral	biological	combination_product	device
441	11	8	107
diagnostic_test	dietary_supplement	drug	genetic
41	11	315	3
other	procedure	radiation	
456	87	8	

## Conditions

Text mining for the most frequently repeated terms in the conditions, as reported in the study registries retrieved from [clinicaltrials.gov](http://clinicaltrials.gov).

## Wordcloud



Figure 1: Wordcloud of most frequently repeated terms in trial registrations

### Chronic vs acute conditions

word	freq
chronic	119
acute	62

### Top 10 terms

word	freq
cancer	208
pain	130
diabetes	103
infections	84
failure	80
depression	74
covid-19	72
dementia	67
breast	66
heart	61

## Package References

- Chang W (2023). *webshot2: Take Screenshots of Web Pages*. R package version 0.1.1, <https://CRAN.R-project.org/package=webshot2>.
- Feinerer I, Hornik K (2024). *tm: Text Mining Package*. R package version 0.7-13, <https://CRAN.R-project.org/package=tm>. Feinerer I, Hornik K, Meyer D (2008). “Text Mining Infrastructure in R.” *Journal of Statistical Software*, 25(5), 1-54. doi:10.18637/jss.v025.i05 <https://doi.org/10.18637/jss.v025.i05>.
- Grolemund G, Wickham H (2011). “Dates and Times Made Easy with lubridate.” *Journal of Statistical Software*, 40(3), 1-25. <https://www.jstatsoft.org/v40/i03/>.
- Hornik K (2020). *NLP: Natural Language Processing Infrastructure*. R package version 0.2-1, <https://CRAN.R-project.org/package=NLP>.
- Iannone R, Cheng J, Schloerke B, Hughes E, Lauer A, Seo J (2024). *gt: Easily Create Presentation-Ready Display Tables*. R package version 0.10.1, <https://CRAN.R-project.org/package=gt>.
- Lang D, Chien G (2018). *wordcloud2: Create Word Cloud by ‘htmlwidget’*. R package version 0.2.1, <https://CRAN.R-project.org/package=wordcloud2>.
- Makowski D, Lüdtke D, Patil I, Thériault R, Ben-Shachar M, Wiernik B (2023). “Automated Results Reporting as a Practical Tool to Improve Reproducibility and Methodological Best Practices Adoption.” *CRAN*. <https://easystats.github.io/report/>.
- Müller K, Wickham H (2023). *tibble: Simple Data Frames*. R package version 3.2.1, <https://CRAN.R-project.org/package=tibble>.
- R Core Team (2024). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.
- Rinker TW, Kurkiewicz D (2018). *pacman: Package Management for R*. version 0.5.0, <http://github.com/trinker/pacman>.
- Vaidyanathan R, Xie Y, Allaire J, Cheng J, Sievert C, Russell K (2023). *htmlwidgets: HTML Widgets for R*. R package version 1.6.4, <https://CRAN.R-project.org/package=htmlwidgets>.
- Wickham H (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. ISBN 978-3-319-24277-4, <https://ggplot2.tidyverse.org>.
- Wickham H (2023). *forcats: Tools for Working with Categorical Variables (Factors)*. R package version 1.0.0, <https://CRAN.R-project.org/package=forcats>.
- Wickham H (2023). *stringr: Simple, Consistent Wrappers for Common String Operations*. R package version 1.5.1, <https://CRAN.R-project.org/package=stringr>.
- Wickham H, Averick M, Bryan J, Chang W, McGowan LD, François R, Grolemund G, Hayes A, Henry L, Hester J, Kuhn M, Pedersen TL, Miller E, Bache SM, Müller K, Ooms J, Robinson D, Seidel DP, Spinu V, Takahashi K, Vaughan D, Wilke C, Woo K, Yutani H (2019). “Welcome to the tidyverse.” *Journal of Open Source Software*, 4(43), 1686. doi:10.21105/joss.01686 <https://doi.org/10.21105/joss.01686>.
- Wickham H, François R, Henry L, Müller K, Vaughan D (2023). *dplyr: A Grammar of Data Manipulation*. R package version 1.1.4, <https://CRAN.R-project.org/package=dplyr>.

[dplyr](#).

- Wickham H, Henry L (2023). *purrr: Functional Programming Tools*. R package version 1.0.2, <https://CRAN.R-project.org/package=purrr>.
- Wickham H, Hester J, Bryan J (2024). *readr: Read Rectangular Text Data*. R package version 2.1.5, <https://CRAN.R-project.org/package=readr>.
- Wickham H, Vaughan D, Girlich M (2024). *tidyr: Tidy Messy Data*. R package version 1.3.1, <https://CRAN.R-project.org/package=tidyr>.