Descriptive Analysis

Indoor dust bacterial and fungal microbiota composition and allergic diseases: a scoping review

Javier Mancilla Galindo, MSc student
Supervisors: Inge Wouters and Alex Bossers
Examiner: Lidwien Smit

2024-06-03

Table of contents

Packages and session information

```
if (!require("pacman", quietly = TRUE)) {
  install.packages("pacman")
}

pacman::p_load(
  tidyverse,    # Used for basic data handling and visualization.
  ggmosaic,     # Used to create mosaic plots.
  RColorBrewer, # Color palettes for data visualization.
  maps,         # Retrieve map data to plot world map.
  gridExtra,    # Arrange grobs in a plot.
  gt,           # Print and save html tables.
  report        # Used to generate package citations in markdown format.
)
```

```
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/London
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          gridExtra_2.3      maps_3.4.2
 [5] RColorBrewer_1.1-3 ggmosaic_0.3.3     lubridate_1.9.3    forcats_1.0.0
 [9] stringr_1.5.1      dplyr_1.1.4        purrr_1.0.2        readr_2.1.5
[13] tidyr_1.3.1        tibble_3.2.1       ggplot2_3.5.1      tidyverse_2.0.0
[17] pacman_0.5.1
```

Load data

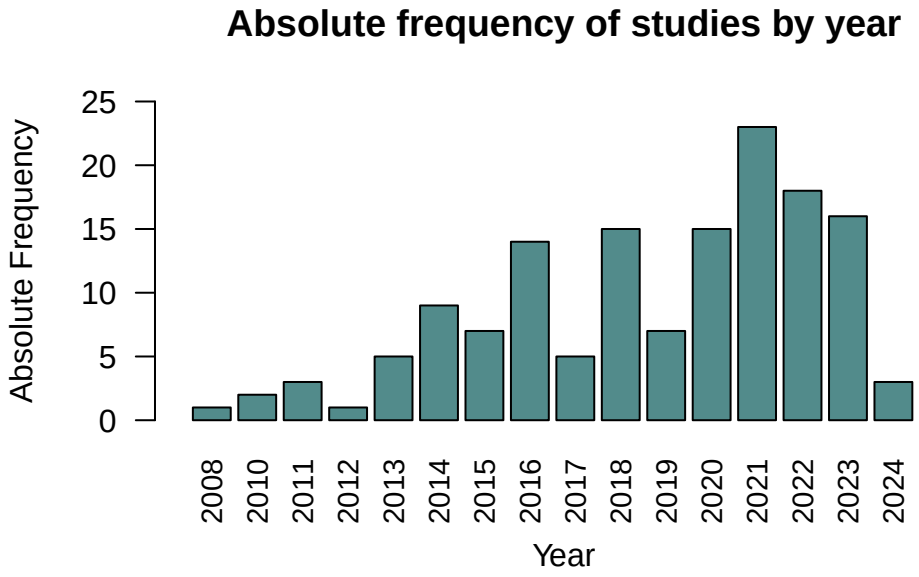
```
Formal class 'DataFrameCollection' [package ".GlobalEnv"] with 0 slots
 Named list()
  ..$ data                        : tibble [144 x 31] (S3: tbl_df/tbl/data.frame)
  ..$ countries                   : tibble [175 x 5] (S3: tbl_df/tbl/data.frame)
  ..$ collectors                  : tibble [184 x 2] (S3: tbl_df/tbl/data.frame)
  ..$ buildings                   : tibble [159 x 2] (S3: tbl_df/tbl/data.frame)
  ..$ environmental_determinants: tibble [595 x 3] (S3: tbl_df/tbl/data.frame)
  ..$ references                  :'data.frame':  144 obs. of  90 variables:
```

Descriptive analysis of variables collected

Year of publication

The range of year of publication was 2008 to 2024.

The following barplot saved as **Figure_years_publication.jpeg** shows the absolute frequency of original research articles published per year.



Microbiome type

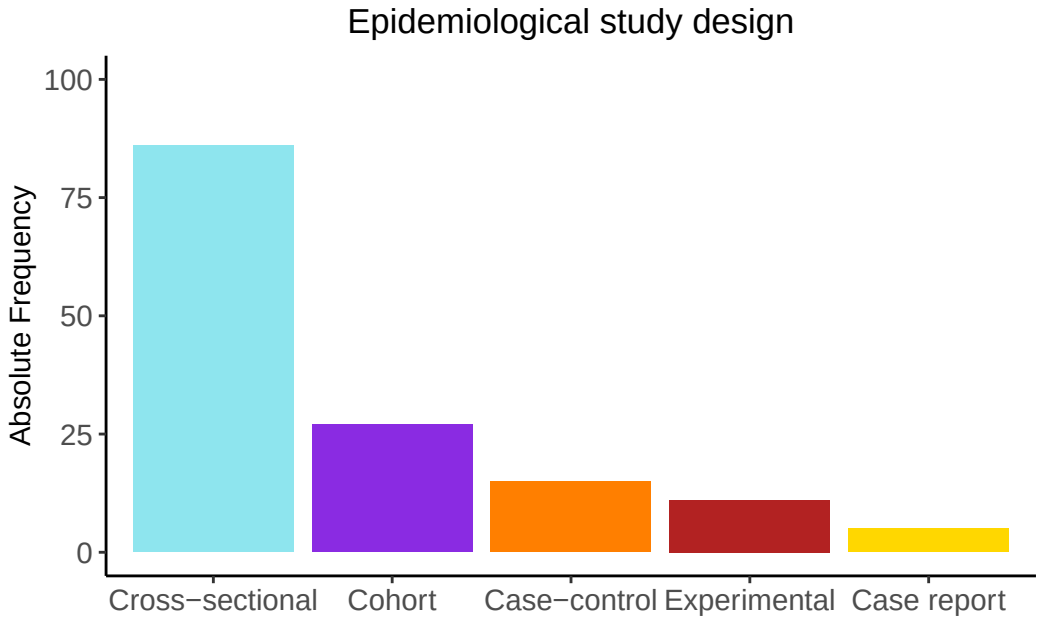
Type of microbiome studied	n	Percentage
Bacterial	60	41.7
Fungal	15	10.4
Bacterial and Fungal	59	41.0
Bacterial, Fungal and other	9	6.2
Other	1	0.7

Available sequence data

Raw sequences provided	n	Percentage
No	60	41.7
Yes	84	58.3

Study Design

Study design	n	Percentage
Cross-sectional	86	59.7
Cohort	27	18.8
Case-control	15	10.4
Experimental	11	7.6
Case report	5	3.5

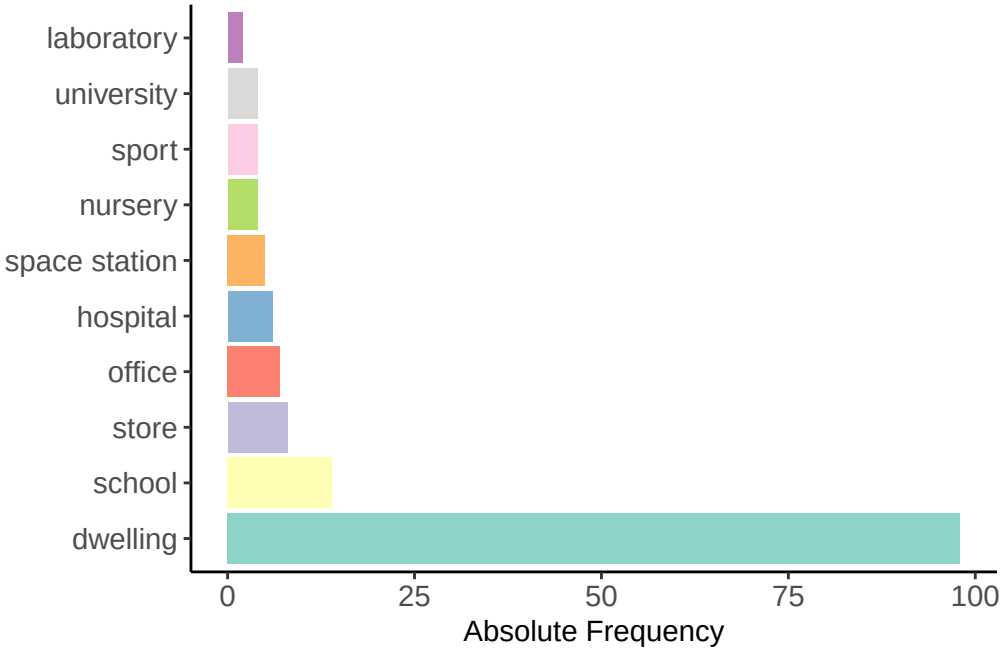


The studies included in the review used the following study designs: Cross-sectional, n=86 (59.7%), Cohort, n=27 (18.8%), Case-control, n=15 (10.4%), Experimental, n=11 (7.6%), Case report, n=5 (3.5%).

Type of building

Type of building	n	Percentage
dwelling	98	68.1
school	14	9.7
store	8	5.6
office	7	4.9
hospital	6	4.2
space station	5	3.5
nursery	4	2.8
sport	4	2.8
university	4	2.8
laboratory	2	1.4
archive	1	0.7
biosphere	1	0.7
church	1	0.7
coal mine	1	0.7
industry	1	0.7
restaurant	1	0.7
vehicle	1	0.7

Barplot top 10 types of building:



Countries

Some of the studies were conducted in multiple countries or in the international space station:

More than 1 country	n	Percentage
No	125	86.8
Yes	14	9.7
ISS	5	3.5

After excluding those studies conducted in the international space station, these are the frequencies and percentages of studies per country.

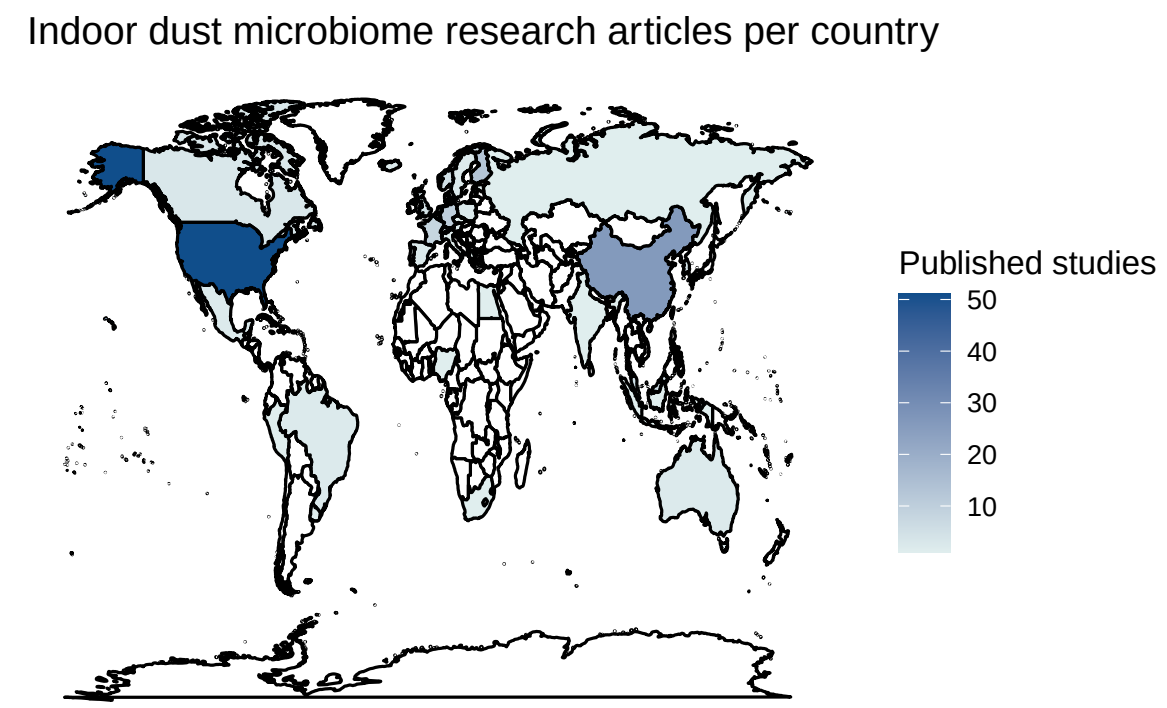
Country of sample collection	n	Percentage
USA	51	29.14
China	26	14.86
Finland	13	7.43
Germany	10	5.71
France	8	4.57
Denmark	6	3.43
Malaysia	6	3.43
Norway	5	2.86
South Korea	4	2.29
Switzerland	4	2.29
United Kingdom	4	2.29
Canada	3	1.71
Australia	2	1.14
Austria	2	1.14
Belgium	2	1.14
Brazil	2	1.14
Iceland	2	1.14
Mexico	2	1.14
Netherlands	2	1.14
Poland	2	1.14
Spain	2	1.14
Sweden	2	1.14
Czech Republic	1	0.57
Egypt	1	0.57
Estonia	1	0.57
Greece	1	0.57
India	1	0.57
Indonesia	1	0.57

Italy	1	0.57
Micronesia	1	0.57
Nigeria	1	0.57
Peru	1	0.57
Portugal	1	0.57
Russia	1	0.57
Singapore	1	0.57
South Africa	1	0.57
Uruguay	1	0.57

Map

I will use the world map data from the maps package `map_data("world")` to join with the `countries` dataset.

```
source("scripts/world_map.R")
```



Region

World region	n	Percentage
Europe & Central Asia	70	40.0
North America	54	30.9
East Asia & Pacific	41	23.4
Latin America & Caribbean	6	3.4
Sub-Saharan Africa	2	1.1
Middle East & North Africa	1	0.6
South Asia	1	0.6

Income group

Income classification	n	Percentage
High income	131	74.9
Upper middle income	40	22.9
Lower middle income	4	2.3

Topics

Allergy

Allergy as topic	n	Percentage
No	103	71.5
Yes	41	28.5

Environmental determinants

Environmental determinants as topic	n	Percentage
No	28	19.4
Yes	116	80.6

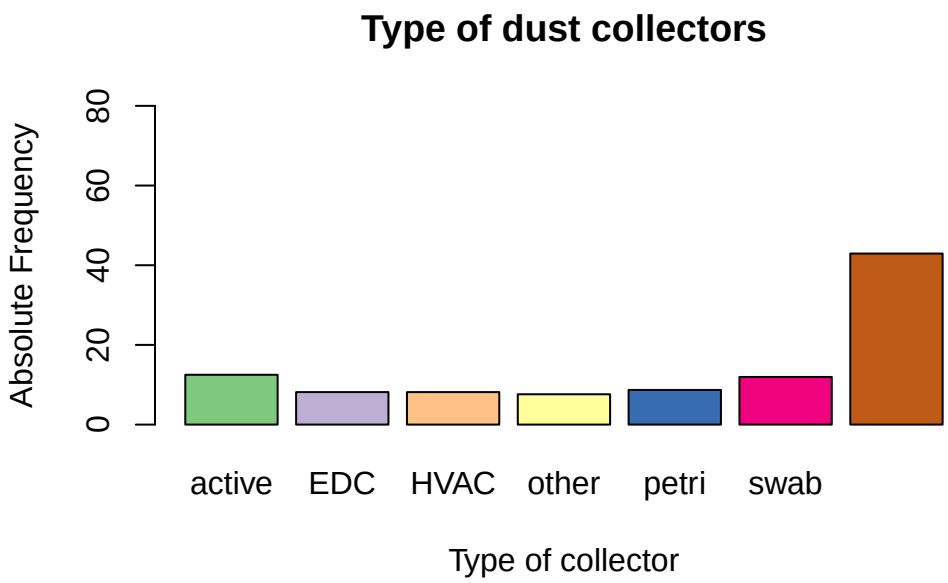
Dust collectors

Multiple collectors

More than 1 collector	n	Percentage
No	115	79.9
Yes	29	20.1

Type of dust collector

Dust sample collector	n	Percentage
vacuum	79	54.9
active-sampler	23	16.0
swab	22	15.3
petri	16	11.1
EDC	15	10.4
HVAC	15	10.4
other	14	9.7



Descriptive vs causal

Causality	n	Percentage
Descriptive	82	56.9
Causal	62	43.1

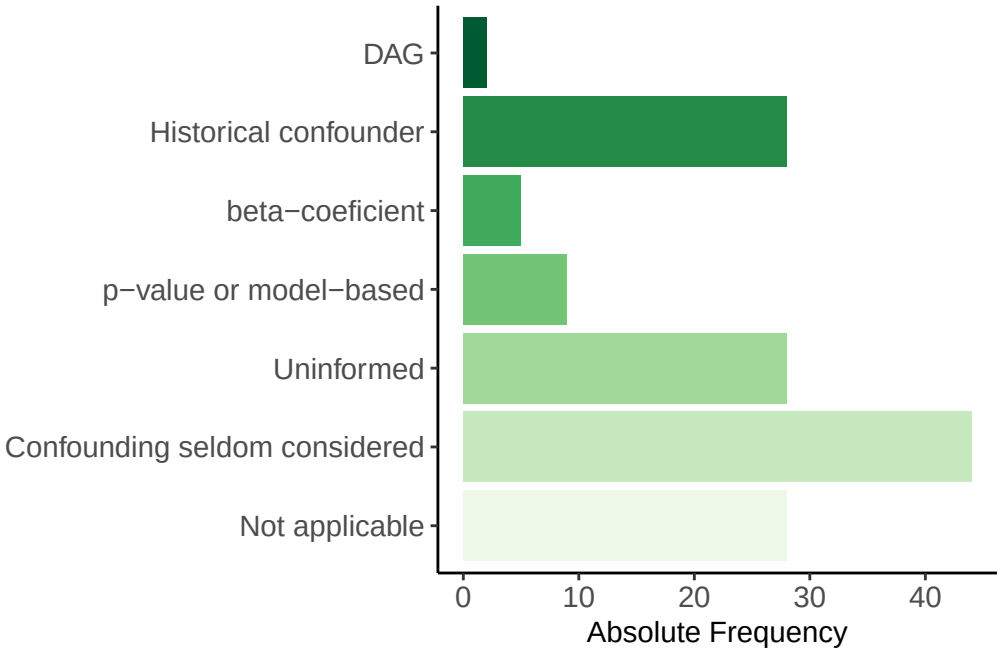
Causal statements

Causal statements	n	Percentage
No	26	18.1
Yes	118	81.9

Confounding

Variable identification method	n	Percentage
Not applicable	28	19.4
Confounding seldom considered	44	30.6
Uninformed	28	19.4
p-value or model-based	9	6.2
beta-coeficient	5	3.5
Historical confounder	28	19.4
DAG	2	1.4

Variable selection



Environmental determinants

Category	n	Percentage
building characteristics	114	79.2
humidity/dampness	56	38.9
pets	51	35.4
air pollutants	42	29.2
building occupants	33	22.9
season	30	20.8
temperature	29	20.1
geography	28	19.4
ventilation	27	18.8
mold	23	16.0
allergen	21	14.6
green environment	19	13.2
urbanicity	19	13.2
smoking	16	11.1
chemicals	14	9.7
other	14	9.7
cleaning habits	13	9.0
outdoor microbiome	12	8.3
farming	10	6.9
infestation	9	6.2
furniture	6	4.2
heating	4	2.8
light	3	2.1
water sources	2	1.4

Table of characteristics of individual studies

The table of characteristics of studies with their citations will be saved as *Table_study_characteristics.csv*.

```
source("scripts/studies_included_table.R")
```

Package references

- Becker OScbRA, Minka ARWRvbRBEbTP, Deckmyn. A (2023). *maps: Draw Geographical Maps*. R package version 3.4.2, https://CRAN.R-project.org/package=maps.
- 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/.
- 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.
- Jeppson H, Hofmann H, Cook D (2021). *ggmosaic: Mosaic Plots in the 'ggplot2' Framework*. R package version 0.3.3, https://CRAN.R-project.org/package=ggmosaic.
- Makowski D, Lüdecke 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.
- Neuwirth E (2022). *RColorBrewer: ColorBrewer Palettes*. R package version 1.1-3, https://CRAN.R-project.org/package=RColorBrewer.
- 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.
- 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.
- 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.