30 Funding Analysis and Grant-to-Publication Linkage

R. Heller · Date: 2026-05-09 · Last changed: 2026-10-31

30.1 Learning objectives

After completing this chapter, you will be able to:

- Extract funder information from OpenAlex work records
- Compute funder-level publication counts and citation impact
- Analyse co-funding patterns (papers acknowledging multiple funders)
- Interpret funding data critically, accounting for coverage and attribution challenges
- Discuss why funding-output linkage is inherently imprecise

30.2 Setup

```
library(tidyverse)
library(openalexR)
library(glue)
library(gt)

set.seed(20260509)

source(here::here("R", "api_helpers.R"))
source(here::here("R", "utils.R"))
source(here::here("R", "funding_helpers.R"))
source(here::here("R", "sci_palette.R"))
```

30.3 Conceptual background

Funding agencies invest billions in research and want to understand what their investment produces. **Funding acknowledgement analysis** extracts funder names and grant identifiers from publication metadata or full text, linking grants to the papers they supported.

OpenAlex includes funder information for many works, derived from Crossref's funding metadata (which in turn comes from publisher-reported FundRef data). Coverage varies: journals that participate in Crossref's FundRef initiative have good coverage; others may have no funding data at all (Priem et al. 2022).

Key analytical questions include: How many publications does a funder support? What is the citation impact of funded vs. unfunded research? Which funders frequently co-fund research? How does a funder's portfolio map onto the landscape of science?

The fundamental challenge is **attribution**: a paper acknowledging three funders and seven authors is the result of contributions from all of them. Assigning full credit to each funder double-counts; fractional credit is more principled but harder to implement. Similar issues arise with multi-source funding: a paper may acknowledge an infrastructure grant, a project grant, and a fellowship, each with different levels of contribution.

30.4 Worked example

30.4.1 Extracting funder data

```
works ← fetch_openalex(
  entity = "works",
  primary_location.source.id = "S148561398",
  from_publication_date = "2020-01-01",
  to_publication_date = "2023-12-31",
  options = list(sample = 400, seed = 42)
)

funding_works ← works ▷
  mutate(funders = map(funders, sci_funder_rows),
         has_funder_metadata = map_int(funders, nrow) > 0)

funded ← funding_works ▷
  filter(has_funder_metadata) ▷
  select(id, display_name, cited_by_count, funders)

cat(glue("Works with observed funder IDs: {nrow(funded)} / {nrow(works)}
({scales::percent(nrow(funded)/nrow(works))})\n"))
```

```
#> Works with observed funder IDs: 228 / 400 (57%)
```

```
if (!nrow(funded)) {
  cat("No usable funder records occur in this sample. Funder summaries are
unavailable; missing metadata does not establish that the papers were unfunded.\n")
}
```

```
funder_data ← if (nrow(funded)) {
  funded ▷
    unnest(funders, names_sep = "_") ▷
    select(work_id = id, funder_id = funders_id,
           funder = funders_display_name, cited_by_count)
} else {
  tibble(work_id = character(), funder_id = character(),
         funder = character(), cited_by_count = numeric())
}

cat(glue("Funder-work pairs: {nrow(funder_data)}\n"))
```

```
#> Funder-work pairs: 365
```

```
cat(glue("Unique funders: {n_distinct(funder_data$funder_id)}\n"))
```

```
#> Unique funders: 195
```

30.4.2 Top funders by publication count

```
funder_counts ← funder_data ▷
  group_by(funder_id) ▷
  summarise(funder = first(funder), n = n_distinct(work_id), .groups = "drop") ▷
  arrange(desc(n))

if (nrow(funder_counts)) {
funder_counts ▷
  head(15) ▷
  mutate(funder = fct_reorder(funder, n)) ▷
  ggplot(aes(x = n, y = funder)) +
  geom_col(fill = palette_sci(1)) +
  labs(x = "Publications supported", y = NULL) +
  theme_sci()
} else {
  cat("No funder publication-count plot: no observed funder records.\n")
}
```

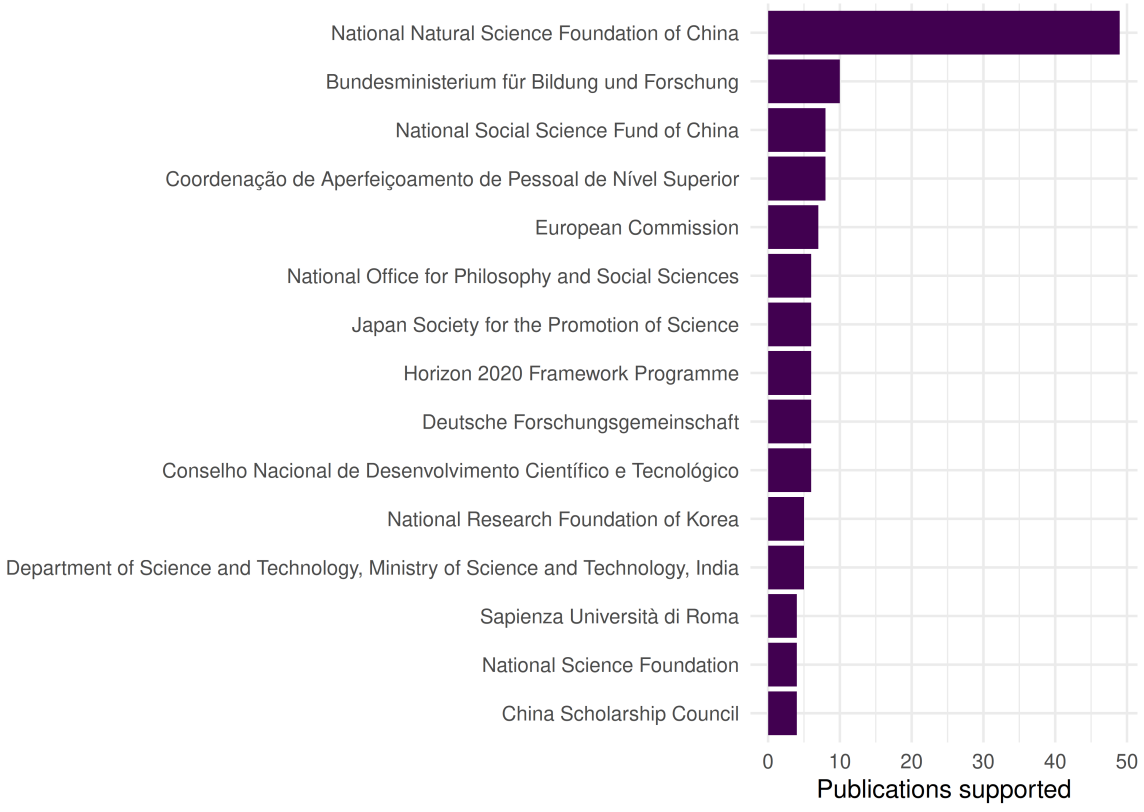


Figure 30.1: Up to 15 funders by the number of sampled publications with observed funder metadata.

30.4.3 Citation impact by funder

```
funder_impact ← funder_data ▷
  group_by(funder_id) ▷
  summarise(
    funder = first(funder),
    n_pubs = n(),
    mean_cites = round(mean(cited_by_count), 1),
    median_cites = median(cited_by_count),
    .groups = "drop"
  ) ▷
  filter(n_pubs ⩾ 5) ▷
  arrange(desc(mean_cites))

if (nrow(funder_impact)) {
  funder_impact ▷ head(10) ▷ gt()
} else {
  cat("No funder has at least five sampled publications with usable metadata; no
funder citation summary is reported.\n")
}
```

funder_id	funder	n_pubs	mean_cites	median_cites
https://openalex.org/F4320322025	Conselho Nacional de Desenvolvimento Científico e Tecnológico	6	58.3	45.5
https://openalex.org/F4320335869	National Social Science Fund of China	8	37.2	20.0
https://openalex.org/F4320327557	National Office for Philosophy and Social Sciences	6	36.0	25.5
https://openalex.org/F4320322120	National Research Foundation of Korea	5	34.2	20.0
https://openalex.org/F4320321091	Coordenação de Aperfeiçoamento de Pessoal de Nível Superior	8	30.8	20.5
https://openalex.org/F4320321114	Bundesministerium für Bildung und Forschung	10	30.8	25.0
https://openalex.org/F4320321001	National Natural Science Foundation of China	49	28.1	24.0
https://openalex.org/F4320320879	Deutsche Forschungsgemeinschaft	6	26.3	25.5

funder_id	funder	n_pubs	mean_cites	median_cites
https://openalex.org/F4320320719	Department of Science and Technology, Ministry of Science and Technology, India	5	26.2	17.0
https://openalex.org/F4320334764	Japan Society for the Promotion of Science	6	24.3	17.0

30.4.4 Comparison by observed funder metadata

```
if (n_distinct(funding_works$has_funder_metadata) == 2L) {
funding_works ▷
  mutate(funder_metadata = if_else(has_funder_metadata,
                                   "Observed funder", "No observed funder")) ▷
  ggplot(aes(x = funder_metadata, y = cited_by_count + 1)) +
  geom_boxplot(fill = palette_sci(1), alpha = 0.7) +
  scale_y_log10() +
  labs(x = "Funder metadata", y = "Citations (log scale)") +
  theme_sci()
} else {
  cat("No two-group citation comparison: only one funder-metadata group is
observed.\n")
}
```

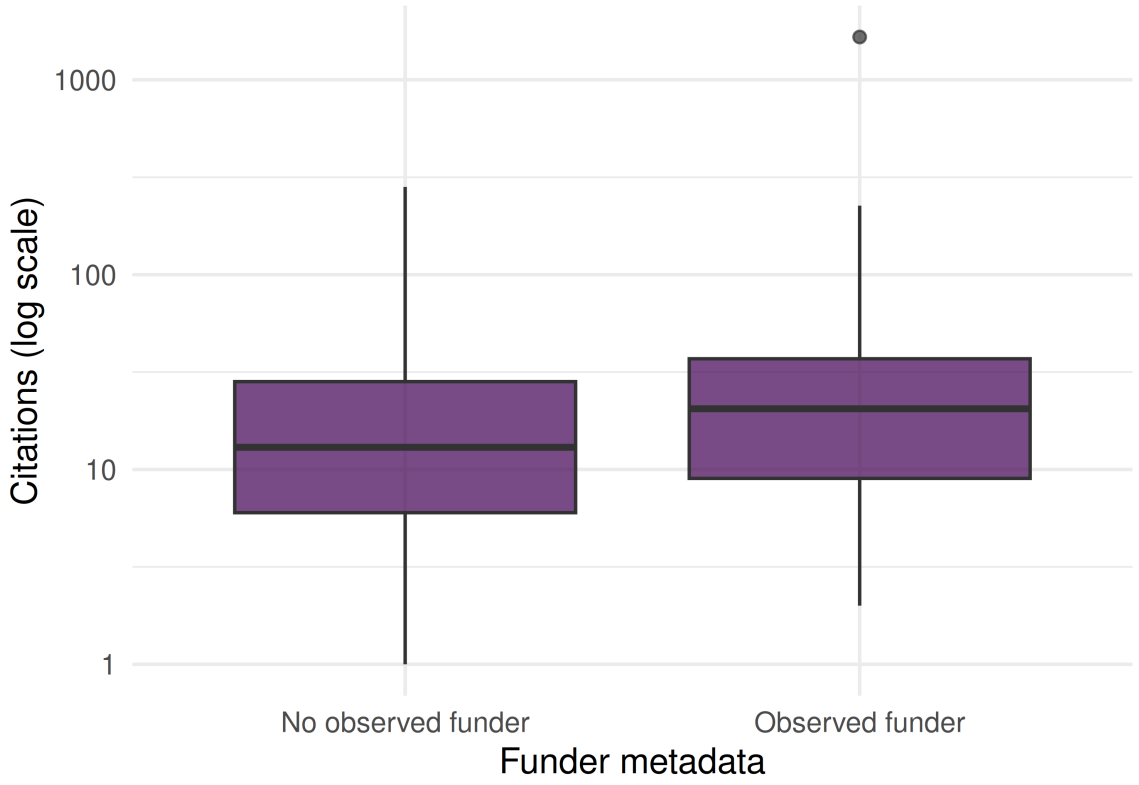


Figure 30.2: Citation distribution by presence of observed funder metadata; missing metadata does not establish absence of funding.

30.5 Diagnostics and interpretation

- **Coverage rate**: The fraction of papers with funder data depends on the journal, publisher, and field. Report this rate before interpreting funder statistics.
- **Funder name normalisation**: The same funder may appear under different names (“NSF”, “National Science Foundation”, “US NSF”). Merge variants using funder IDs where available.
- **Selection and metadata coverage**: Papers with observed funder records may differ systematically from papers without them. Their citation gap can reflect selection and publisher reporting; it does not identify the causal effect of funding or compare confirmed funded and unfunded groups.
- **Multi-funder attribution**: Papers acknowledging multiple funders inflate the publication count for each. Consider fractional counting.

30.6 Limitations and responsible use

30.7 Limitations and responsible use

- **Coverage is incomplete.** Not all publishers report funding data to Crossref. Missing funder acknowledgements do not mean research was unfunded.
- **Attribution is ambiguous.** A grant acknowledgement does not specify which part of the work the grant supported. Infrastructure grants, salary support, and project funding all appear identically.
- **Correlation ≠ causation.** Funded papers receive more citations, but this may reflect that better-resourced labs produce both more funding applications and more cited papers.
- **Do not rank funders by citation impact.** Funders support different fields, career stages, and risk levels. Comparing mean citations across funders is misleading without field normalisation (Hicks et al. 2015).

30.8 Common pitfalls

30.9 Common pitfalls

- **Ignoring funder coverage rates.** Comparing funder portfolios without accounting for differential coverage produces biased results.
- **Treating every acknowledgement as equal.** A paper may acknowledge a major project grant and a minor travel grant. Both appear identically in metadata.
- **Comparing across fields.** A biomedical funder's publications will have higher citation counts than a humanities funder's publications due to field-level citation norms.
- **Using funder data for individual evaluation.** Whether a researcher has funding reflects many factors beyond research quality: field norms, career stage, institutional support, and luck.

30.10 Exercises

1. **Funder co-occurrence.** Identify papers with multiple funders. Which funder pairs most frequently co-fund research?

2. **Funder portfolios.** For two major funders, fetch their supported publications and compare topical profiles using keyword analysis.

3. **Temporal trends.** Track the number of papers with funding acknowledgements by year. Is the coverage rate increasing over time?

30.11 Solutions

Solutions are provided in E.

30.12 Further reading

- Priem et al. (2022) — OpenAlex funder metadata from Crossref FundRef.
- Hicks et al. (2015) — Responsible use of metrics in funding evaluation.
- Waltman (2016) — Citation indicators and their application to funding assessment.

30.13 Session info

```
sessionInfo()
```

```
#> R version 4.6.1 (2026-06-24)
#> Platform: x86_64-pc-linux-gnu
#> Running under: Ubuntu 24.04.5 LTS
#>
#> Matrix products: default
#> BLAS:   /usr/lib/x86_64-linux-gnu/openblas-pthread/libblas.so.3
#> LAPACK: /usr/lib/x86_64-linux-gnu/openblas-pthread/libopenblasp-r0.3.26.so;  LAPACK
version 3.12.0
#>
#> locale:
#>  [1] LC_CTYPE=C.UTF-8       LC_NUMERIC=C           LC_TIME=C.UTF-8
#>  [4] LC_COLLATE=C.UTF-8     LC_MONETARY=C.UTF-8    LC_MESSAGES=C.UTF-8
#>  [7] LC_PAPER=C.UTF-8       LC_NAME=C              LC_ADDRESS=C
#> [10] LC_TELEPHONE=C         LC_MEASUREMENT=C.UTF-8 LC_IDENTIFICATION=C
#>
#> time zone: UTC
#> tzcode source: system (glibc)
#>
#> attached base packages:
#> [1] stats     graphics  grDevices utils     datasets  methods   base
#>
#> other attached packages:
#>  [1] uwot_0.2.5                Matrix_1.7-5
#>  [3] word2vec_0.4.1            stm_1.3.10
#>  [5] topicmodels_0.2-17        quanteda.textstats_0.97.2
#>  [7] visNetwork_2.1.4          ggraph_2.2.2
#>  [9] tidygraph_1.3.1           igraph_2.3.4
#> [11] quanteda_4.5.1            pdftools_3.9.1
#> [13] arrow_26.0.1.1            bibliometrix_5.5.0
#> [15] RefManageR_1.4.0          bib2df_1.1.2.0
#> [17] rcrossref_1.2.1           gt_1.4.0
#> [19] tidytext_0.4.3            glue_1.8.1
#> [21] openalexR_3.1.2           lubridate_1.9.5
#> [23] forcats_1.0.1             stringr_1.6.0
#> [25] dplyr_1.2.1               purrr_1.2.2
#> [27] readr_2.2.0               tidyr_1.3.2
#> [29] tibble_3.3.1              ggplot2_4.0.4
#> [31] tidyverse_2.0.0
#>
#> loaded via a namespace (and not attached):
#>   [1] bibtex_0.5.3           RColorBrewer_1.1-3     rstudioapi_0.19.0
#>   [4] jsonlite_2.0.0         magrittr_2.0.5         modeltools_0.2-25
#>   [7] farver_2.1.2           rmarkdown_2.32         fs_2.1.0
#>  [10] vctrs_0.7.3            memoise_2.0.1          askpass_1.2.1
#>  [13] base64enc_0.1-6        htmltools_0.5.9        contentanalysis_1.1.1
#>  [16] curl_8.1.0             janeaustenr_1.0.0      cellranger_1.1.0
#>  [19] sass_0.5.0             bslib_0.12.0           htmlwidgets_1.6.4
#>  [22] tokenizers_0.3.0       plyr_1.8.9             httr2_1.3.0
#>  [25] plotly_4.12.1          cachem_1.1.0           dimensionsR_0.0.3
#>  [28] mime_0.13              lifecycle_1.0.5        pkgconfig_2.0.3
#>  [31] R6_2.6.1               fastmap_1.2.0          shiny_1.15.0
#>  [34] digest_0.6.39          patchwork_1.3.2        shinycssloaders_1.1.0
#>  [37] rprojroot_2.1.1        RSpectra_0.16-2        SnowballC_0.7.1
#>  [40] labeling_0.4.3         urltools_1.7.3.1       timechange_0.4.0
#>  [43] mgcv_1.9-4             polyclip_1.10-7        httr_1.4.9
#>  [46] compiler_4.6.1         here_1.0.2             bit64_4.8.6
#>  [49] withr_3.0.3            S7_1.0.0               viridis_0.6.5
#>  [52] ggforce_0.5.0          MASS_7.3-65            bibliometrixData_0.3.0
#>  [55] tools_4.6.1            otel_0.3.0             stopwords_2.3
#>  [58] zip_3.0.2              httpuv_1.6.17          rentrez_1.2.4
#>  [61] nlme_3.1-169           promises_1.5.0         grid_4.6.1
#>  [64] stringdist_0.9.17      reshape2_1.4.5         generics_0.1.4
#>  [67] gtable_0.3.6           tzdb_0.5.0             rscopus_0.9.0
```

```
#>  [70] ca_0.72                data.table_1.18.6.1    hms_1.1.4
#>  [73] xml2_1.6.0             utf8_1.2.6             ggrepel_0.9.8
#>  [76] pillar_1.11.1          nsyllable_1.0.1        vroom_1.7.1
#>  [79] later_1.4.8            splines_4.6.1          tweenr_2.0.3
#>  [82] lattice_0.22-9         FNN_1.1.4.1            bit_4.6.0
#>  [85] tidyselect_1.2.1       tm_0.7-20              miniUI_0.1.2
#>  [88] downlit_0.4.5          knitr_1.52             gridExtra_2.3.1
#>  [91] NLP_0.3-3              bookdown_0.48          stats4_4.6.1
#>  [94] crul_1.6.0             xfun_0.62              graphlayouts_1.2.5
#>  [97] matrixStats_1.5.0      DT_0.34.0              humaniformat_0.6.0
#> [100] stringi_1.8.9          qpdf_1.4.1             yaml_2.3.12
#> [103] evaluate_1.0.5         httpcode_0.3.0         cli_3.6.6
#> [106] xtable_1.8-8           jquerylib_0.1.4        Rcpp_1.1.2
#> [109] readxl_1.6.0           triebeard_0.4.1        XML_3.99-0.25
#> [112] parallel_4.6.1         assertthat_0.2.1       pubmedR_1.1.0
#> [115] slam_0.1-56            viridisLite_0.4.3      scales_1.4.0
#> [118] crayon_1.5.3           openxlsx_4.2.9         rlang_1.3.0
#> [121] fastmatch_1.1-8
```