14 Temporal Analysis and Citation Aging

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

14.1 Learning objectives

After completing this chapter, you will be able to:

- Define and apply citation windows of different lengths
- Compute and interpret the citation half-life for a journal or corpus
- Plot citation aging curves showing how citations accumulate over time
- Calculate the Price index (proportion of references less than five years old)
- Analyse temporal trends in publication output and citation rates
- Explain how aging patterns differ across disciplines

14.2 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", "sci_palette.R"))
```

14.3 Conceptual background

Citations are not evenly distributed over time. Most papers receive the bulk of their citations within a few years of publication, then gradually fade. This phenomenon — **citation aging** or **obsolescence** — was first studied systematically by Solla Price (1963), who observed that scientific literature has a characteristic "half-life": the time by which half of all citations to a given year's output have been received.

The **citation half-life** varies dramatically across disciplines. In fast-moving fields like molecular biology, half-lives may be three to four years; in mathematics or humanities, they can exceed ten years. This has direct consequences for metrics: a two-year Journal Impact Factor systematically undervalues journals in slow-aging fields (Waltman 2016).

A **citation window** is the fixed period after publication during which citations are counted. The JIF uses a two-year window; CiteScore uses four years. Choosing the window involves a trade-off: short windows favour fast-aging fields but miss delayed-recognition papers; long windows are more inclusive but dilute the signal of recent trends.

The **Price index** — the proportion of references in a paper's bibliography that are less than five years old — measures how rapidly a field consumes recent literature. A high Price index indicates a fast-moving frontier field; a low Price index suggests reliance on older foundational works. Solla Price (1963) introduced this as an "immediacy" measure.

Sleeping beauties — papers that receive few citations for years then suddenly attract attention (Kleinberg 2003) — violate the standard aging model. Their existence is a reminder that citation half-lives describe averages, not individual papers. Garfield (1955) already noted that some ideas take decades to be recognised.

Understanding citation temporality is essential for fair evaluation. Comparing citation counts without accounting for publication year, citation window, and field-specific aging rates produces meaningless results.

14.4 Worked example

14.4.1 Acquiring the data

We fetch works from two journals with different expected aging profiles.

```
# Scientometrics (social science, moderate aging)
works_sciento ← fetch_openalex(
  entity = "works",
  primary_location.source.id = "S148561398",
  from_publication_date = "2013-01-01",
  to_publication_date = "2018-12-31",
  type = "article",
  options = list(sample = 400, seed = 42)
)

# PLOS ONE (multidisciplinary, faster aging)
works_plos ← fetch_openalex(
  entity = "works",
  primary_location.source.id = "S202381698",
  from_publication_date = "2013-01-01",
  to_publication_date = "2018-12-31",
  type = "article",
  options = list(sample = 400, seed = 42)
)

works_all ← bind_rows(
  works_sciento ▷ mutate(journal = "Scientometrics"),
  works_plos ▷ mutate(journal = "PLOS ONE")
) ▷
  mutate(year = year(publication_date),
         age = 2024 - year)
```

14.4.2 Citation aging curves

We compute mean citations by age (years since publication) for each journal.

```r
aging ← works_all ▷
  group_by(journal, age) ▷
  summarise(
    mean_cites = mean(cited_by_count, na.rm = TRUE),
    median_cites = median(cited_by_count, na.rm = TRUE),
    n = n(),
    .groups = "drop"
  ) ▷
  arrange(journal, age)
```

```r
ggplot(aging, aes(x = age, y = mean_cites, colour = journal)) +
  geom_line(linewidth = 1) +
  geom_point(size = 2) +
  scale_colour_manual(values = palette_sci(2)) +
  labs(x = "Years since publication", y = "Mean citations",
       colour = "Journal") +
  theme_sci()
```

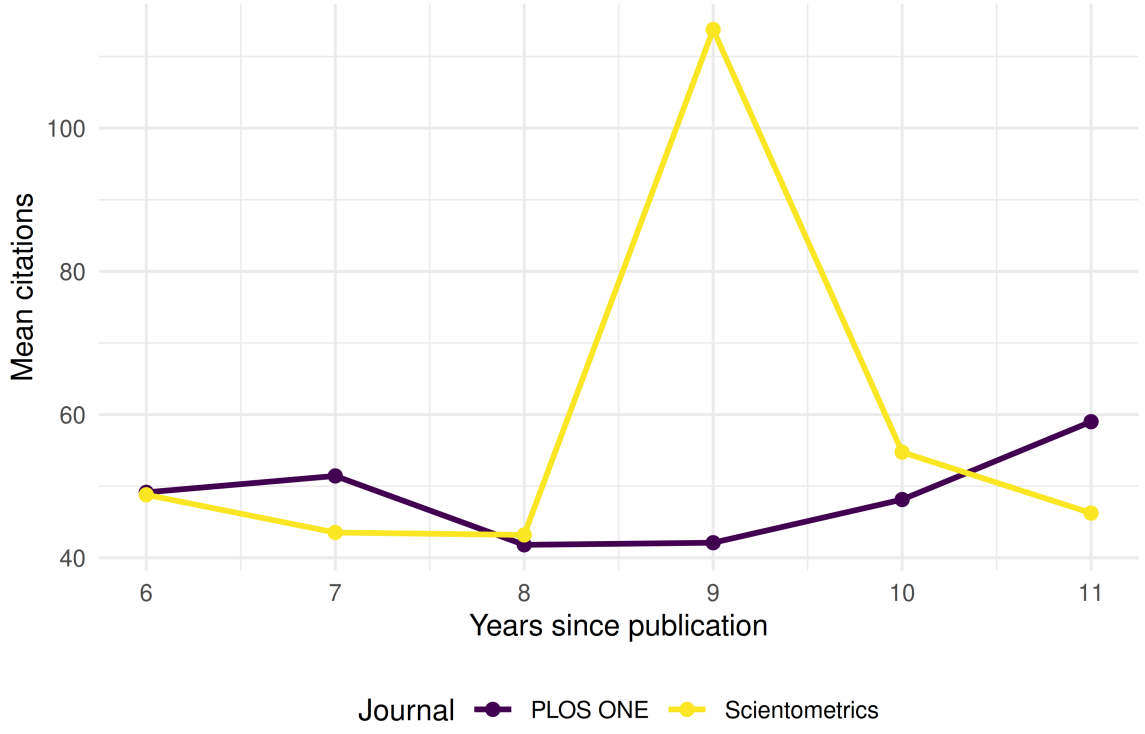


Figure 14.1: Citation aging curves for Scientometrics and PLOS ONE articles published 2013-2018.

14.4.3 Estimating citation half-life

We estimate the half-life as the age at which cumulative citations reach 50% of total.

```
half_life ← works_all ▷
  group_by(journal) ▷
  arrange(age) ▷
  mutate(
    cum_cites = cumsum(cited_by_count),
    total_cites = sum(cited_by_count),
    cum_pct = cum_cites / total_cites
  ) ▷
  summarise(
    total_cites = first(total_cites),
    half_life_age = min(age[cum_pct ⩾ 0.50]),
    .groups = "drop"
  )

half_life ▷
  gt() ▷
  cols_label(
    journal = "Journal",
    total_cites = "Total citations",
    half_life_age = "Half-life (years)"
  )
```

Journal	Total citations	Half-life (years)
PLOS ONE	19581	9
Scientometrics	22651	9

14.4.4 Price index

The Price index is the fraction of references less than five years old. A full computation requires publication years of each reference; here we show reference-list lengths as a starting point.

```
works_all ▷
  filter(!is.na(referenced_works_count), referenced_works_count > 0) ▷
  group_by(journal) ▷
  summarise(mean_refs = round(mean(referenced_works_count), 1), .groups = "drop")
```

```
#> # A tibble: 2 × 2
#>   journal        mean_refs
#>   <chr>              <dbl>
#> 1 PLOS ONE            47.1
#> 2 Scientometrics      42.3
```

14.4.5 Cumulative citation trajectories

```
cum_aging ← works_all ▷
  group_by(journal, age) ▷
  summarise(total = sum(cited_by_count), .groups = "drop") ▷
  group_by(journal) ▷
  arrange(age) ▷
  mutate(
    cum_total = cumsum(total),
    cum_pct = cum_total / sum(total)
  )

ggplot(cum_aging, aes(x = age, y = cum_pct, colour = journal)) +
  geom_line(linewidth = 1) +
  geom_hline(yintercept = 0.5, linetype = "dashed", colour = "grey50") +
  scale_y_continuous(labels = scales::percent) +
  scale_colour_manual(values = palette_sci(2)) +
  labs(x = "Years since publication", y = "Cumulative % of citations",
       colour = "Journal") +
  annotate("text", x = 9, y = 0.52, label = "50% threshold",
           colour = "grey40", size = 3.5) +
  theme_sci()
```

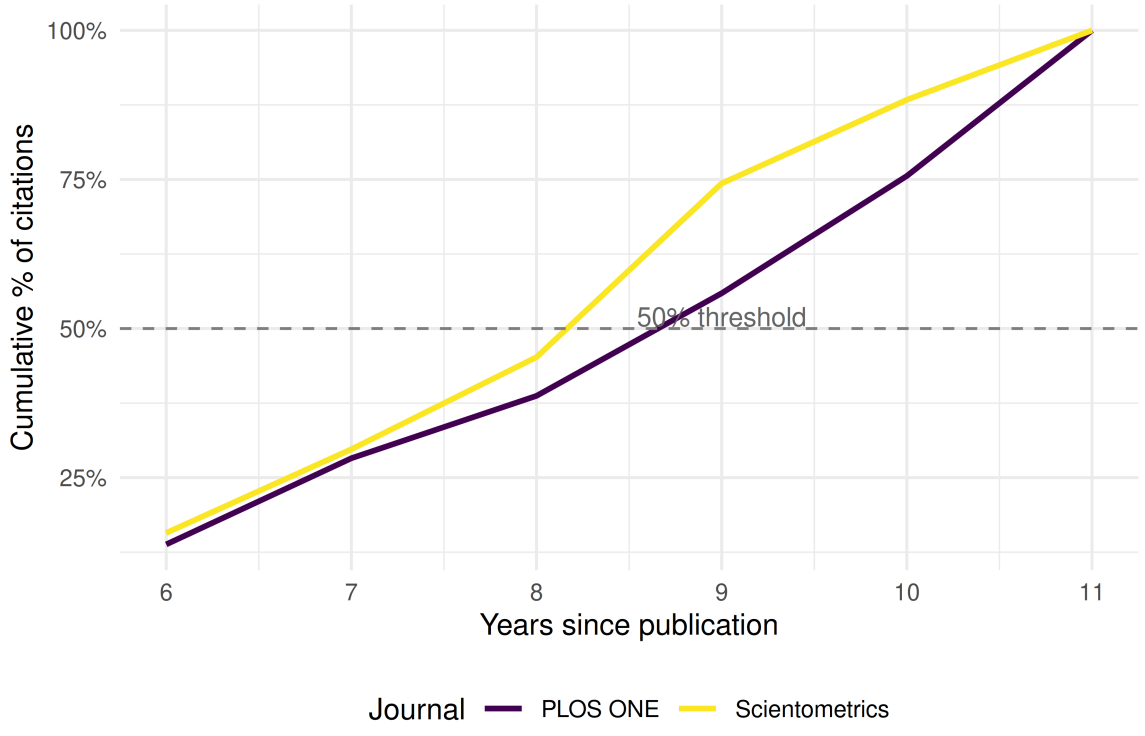


Figure 14.2: Cumulative citation percentage by years since publication.

14.5 Diagnostics and interpretation

- **Sample size per age bin**: Ensure each year has enough papers for stable estimates. Small cohorts produce noisy aging curves.

- **Truncation**: Papers published in the most recent year have not had time to accumulate citations. Exclude or flag incomplete citation windows.
- **Disciplinary benchmarks**: Compare your half-life estimates against known benchmarks. Social sciences typically have half-lives of 5-8 years; biomedical fields 3-5 years; mathematics 8-12 years.
- **Outliers**: A few extremely highly cited papers can distort mean-based aging curves. Compare mean and median trajectories.

14.6 Limitations and responsible use

14.7 Limitations and responsible use

- **Half-life is an average.** Individual papers deviate wildly from the average aging pattern. Sleeping beauties and flash-in-the-pan papers both exist.
- **Database lag.** Citation counts in OpenAlex may lag behind reality by weeks or months. Very recent citation data is unreliable.
- **Survivor bias.** Retracted papers or papers removed from databases disappear from aging analyses, potentially biasing half-life estimates.
- **Citation inflation.** The total volume of citations has grown over time. A paper published in 2020 will accumulate citations faster than an identical paper published in 2000, simply because there are more papers being published (Solla Price 1963).
- **Do not penalise slow-aging fields.** Using short citation windows systematically undervalues mathematics, humanities, and theoretical disciplines (Hicks et al. 2015).

14.8 Common pitfalls

14.9 Common pitfalls

- **Mixing citation windows.** Comparing a 2-year JIF for one journal with lifetime citations for another produces meaningless rankings.
- **Ignoring publication-year effects.** A corpus spanning 2000-2023 includes papers with 23 years of citation history alongside papers with 1 year. Normalise by publication year.
- **Confusing obsolescence with irrelevance.** A declining citation curve does not mean the knowledge is wrong or useless; it may mean the ideas have been absorbed into textbooks or standard practice.

- **Forgetting about the denominator.** If the number of publications in a field is growing, the absolute citation count grows even if the per-paper impact is constant.

14.10 Exercises

1. **Half-life comparison.** Fetch works from a mathematics journal and a biomedical journal. Compute and compare their citation half-lives. Does the difference match disciplinary expectations?
2. **Price index computation.** For a set of papers with known reference lists, compute the actual Price index (fraction of references < 5 years old). *(Hint: you will need to fetch metadata for the referenced works.)*
3. **Sleeping beauty detection.** Identify papers that received fewer than 2 citations in their first 3 years but more than 20 total. What are these papers about?
4. **Median vs. mean aging.** Replot the aging curves using median instead of mean citations. How does the shape change, and why?
5. **Publication-cohort sensitivity.** Using the fetched Scientometrics sample, compare current cumulative citation summaries for the 2-, 3- and 5-year publication cohorts ending in 2018. Report observed denominators and explain why changing the publication cohort does not establish a citation-receipt window or compute JIF.

14.11 Solutions

Solutions are provided in E.

14.12 Further reading

- Solla Price (1963) — *Little Science, Big Science*; introduced citation aging and the Price index.
- Garfield (1955) — The origin of citation indexing, including early observations on citation temporality.
- Kleinberg (2003) — Bursty patterns in temporal data; relevant to sleeping beauties and sudden citation surges.
- Waltman (2016) — Review of citation indicators, including discussion of citation windows and aging.
- Hicks et al. (2015) — The Leiden Manifesto; warns against penalising slow-aging fields with short citation windows.

14.13 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] quanteda_4.5.1     pdftools_3.9.1     arrow_26.0.1.1     bibliometrix_5.5.0
#>  [5] RefManageR_1.4.0   bib2df_1.1.2.0     rcrossref_1.2.1    gt_1.4.0
#>  [9] tidytext_0.4.3     glue_1.8.1         openalexR_3.1.2    lubridate_1.9.5
#> [13] forcats_1.0.1      stringr_1.6.0      dplyr_1.2.1        purrr_1.2.2
#> [17] readr_2.2.0        tidyr_1.3.2        tibble_3.3.1       ggplot2_4.0.4
#> [21] tidyverse_2.0.0
#>
#> loaded via a namespace (and not attached):
#>   [1] gridExtra_2.3.1        httr2_1.3.0            readxl_1.6.0
#>   [4] rlang_1.3.0            magrittr_2.0.5         otel_0.3.0
#>   [7] compiler_4.6.1         vctrs_0.7.3            httpcode_0.3.0
#>  [10] pkgconfig_2.0.3        fastmap_1.2.0          labeling_0.4.3
#>  [13] ca_0.72                utf8_1.2.6             promises_1.5.0
#>  [16] rmarkdown_2.32         tzdb_0.5.0             bit_4.6.0
#>  [19] xfun_0.62              cachem_1.1.0           jsonlite_2.0.0
#>  [22] SnowballC_0.7.1        later_1.4.8            parallel_4.6.1
#>  [25] stopwords_2.3          R6_2.6.1               bslib_0.12.0
#>  [28] stringi_1.8.9          RColorBrewer_1.1-3     cellranger_1.1.0
#>  [31] jquerylib_0.1.4        assertthat_0.2.1       Rcpp_1.1.2
#>  [34] bookdown_0.48          knitr_1.52             triebeard_0.4.1
#>  [37] base64enc_0.1-6        rentrez_1.2.4          httpuv_1.6.17
#>  [40] Matrix_1.7-5           igraph_2.3.4           timechange_0.4.0
#>  [43] tidyselect_1.2.1       stringdist_0.9.17      pubmedR_1.1.0
#>  [46] rstudioapi_0.19.0      yaml_2.3.12            viridis_0.6.5
#>  [49] miniUI_0.1.2           humaniformat_0.6.0     curl_8.1.0
#>  [52] qpdf_1.4.1             lattice_0.22-9         plyr_1.8.9
#>  [55] shiny_1.15.0           withr_3.0.3            S7_1.0.0
#>  [58] askpass_1.2.1          evaluate_1.0.5         zip_3.0.2
#>  [61] xml2_1.6.0             shinycssloaders_1.1.0  pillar_1.11.1
#>  [64] janeaustenr_1.0.0      DT_0.34.0              plotly_4.12.1
#>  [67] generics_0.1.4         rprojroot_2.1.1        hms_1.1.4
#>  [70] scales_1.4.0           xtable_1.8-8           contentanalysis_1.1.1
#>  [73] tools_4.6.1            data.table_1.18.6.1    tokenizers_0.3.0
#>  [76] openxlsx_4.2.9         XML_3.99-0.25          fs_2.1.0
#>  [79] visNetwork_2.1.4       fastmatch_1.1-8        grid_4.6.1
#>  [82] bibtex_0.5.3           urltools_1.7.3.1       rscopus_0.9.0
#>  [85] dimensionsR_0.0.3      bibliometrixData_0.3.0 cli_3.6.6
#>  [88] viridisLite_0.4.3      downlit_0.4.5          gtable_0.3.6
#>  [91] sass_0.5.0             digest_0.6.39          ggrepel_0.9.8
#>  [94] crul_1.6.0             htmlwidgets_1.6.4      farver_2.1.2
#>  [97] memoise_2.0.1          htmltools_0.5.9        lifecycle_1.0.5
```

```
#> [100] httr_1.4.9             here_1.0.2             mime_0.13
#> [103] bit64_4.8.6
```