



# ABCal: a Python package for author bias computation and scientometric plotting for reviews and meta-analyses

Louis-Stéphane Le Clercq

## ► To cite this version:

Louis-Stéphane Le Clercq. ABCal: a Python package for author bias computation and scientometric plotting for reviews and meta-analyses. *Scientometrics*, 2023, 129, pp.581-600. 10.1007/s11192-023-04880-6 . hal-04308276

**HAL Id: hal-04308276**

**<https://hal.science/hal-04308276>**

Submitted on 27 Nov 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License



# ABCal: a Python package for author bias computation and scientometric plotting for reviews and meta-analyses

Louis-Stéphane Le Clercq<sup>1,2</sup> 

Received: 4 August 2023 / Accepted: 10 November 2023  
© The Author(s) 2023

## Abstract

Systematic reviews are critical summaries of the exiting literature on a given subject and, when combined with meta-analysis, provides a quantitative synthesis of evidence to direct and inform future research. Such reviews must, however, account for complex sources of between study heterogeneity and possible sources of bias, such as publication bias. This paper presents the methods and results of a research study using a newly developed software tool called ABCal (version 1.0.2) to compute and assess author bias in the literature, providing a quantitative measure for the possible effect of overrepresented authors introducing bias to the overall interpretation of the literature. ABCal includes a new metric referred to as author bias, which is a measure of potential biases per paper when the frequency or proportions of contributions from specific authors are considered. The metric is able to account for a significant portion of the observed heterogeneity between studies included in meta-analyses. A meta-regression between observed effect measures and author bias values revealed that higher levels of author bias were associated with higher effect measures while lower author bias was evident for studies with lower effect measures. Furthermore, the software's capabilities to analyse authorship contributions and produce scientometric plots was able to reveal distinct patterns in both the temporal and geographic distributions of publications, which may relate to any evident publication bias. Thus, ABCal can aid researchers in gaining a deeper understanding of the research landscape and assist in identifying both key contributors and holistic research trends.

**Keywords** Reviews · Authorship · Bias detection · Author bias · Scientometric · Plots

## Introduction

Scientific studies based on the empirical method contribute dozens of new publications pertaining to a central hypothesis under investigation on an annual basis. Over time, however, it may become apparent that the evidence in support of, or opposing, fundamental views within a discipline may not be unanimous (Le Clercq et al., [2023a](#), [2023b](#), [2023c](#))

---

✉ Louis-Stéphane Le Clercq  
leclercq.l.s@gmail.com

<sup>1</sup> South African National Biodiversity Institute, Pretoria 0001, South Africa

<sup>2</sup> Department of Genetics, University of the Free State, Bloemfontein 9300, South Africa

and can confound the overall interpretation of primary findings. As existing evidence serves as the foundation for informing the direction of prospective studies, however, scientists are often faced with the difficult task of reading and interpreting the literature to derive central tenets for their subject or discipline (Boell & Cecez-Kecmanovic, 2010; Webster & Watson, 2002). One convenient method to synthesise and assess the existing evidence is through systematic review and meta-analysis.

Systematic reviews were first developed as a tool for the synthesis of evidence for causality (Le Clercq et al., 2016) or treatment (Honvo et al., 2019) in medical research and can be defined as: “a review using a systematic method to summarize evidence on questions with a detailed and comprehensive plan of study” (Tawfik et al., 2019). Thus, a systematic review seeks to identify and critically evaluate all those studies (Dickersin et al., 1994) pertaining to a specific research question for the purposes of deriving conclusions from the full body of evidence rather than relying on individual studies alone. Furthermore, systematic reviews attempt to standardise the methods (Moher et al., 2010; O’Dea et al., 2021) used to identify and screen studies in a way that is comprehensive, transparent, and above all reproducible and can serve as independent studies (Kraus et al., 2022). This avoids some of the pitfalls and biases that could influence narrative reviews (Pae, 2015; Tawfik et al., 2019). Another advantage of systematic reviews is the possibility to perform a meta-analysis and scientometric assessment of the included studies (Nakagawa et al., 2023). This is done using primary reported statistics to derive the effect size (Cohen, 1988) or treatment effect (TE), and variance or standard error (SETE), of the measured outcome. This facilitates between-study comparisons and enables the calculation of a pooled effect through a fixed- or random effects model (Borenstein et al., 2010). The pooled effect, therefor, serves as a quantitative measure of the total evidence.

Meta-analysis is not without possible confounders: several factors could contribute to between study differences, called heterogeneity (Higgins & Thompson, 2002), or be a source of bias (Felson, 1992; Sterne et al., 2001). Heterogeneity is expressed by two statistics, the heterogeneity measure ( $I^2$ ) and the between-study variance or tau-squared ( $\tau^2$ ). The  $I^2$  measure expresses the percentage of total variance in the effect sizes that is explained by between-study variance. The  $\tau^2$  approximates between-study variances but is reliant upon the specific effect sizes and needs to be quantitated by a P-value (Higgins, 2008). As heterogeneity could potentially reduce the ability to compare or combine the outcomes from all studies that meet inclusion criteria, it is critical for authors to identify possible sources of heterogeneity and attempt to account for them.

Common factors that contribute to heterogeneity include sample size, quantitative method, and study population. Two approaches can be used to account for these variables: meta-analysis with subgroups and meta-regression of factors. The first, meta-analysis with subgroups, determines if effect sizes and their corresponding variances differ between subgroups (Borenstein & Higgins, 2013). In the case of study populations representing different species, as may be the case in reviews in animal sciences or ecology, several methods have also been developed to account for phylogeny (Chamberlain et al., 2012) and taxonomy by performing a phylogenetic meta-analysis (Adams, 2008; Lajeunesse, 2009). The second approach, meta regression, determines if a significant part of the heterogeneity can be accounted for by individual study attributes, which may be more useful for continuous variables such as sample size (Baker et al., 2009).

The other, and perhaps more difficult, task is to identify and quantify potential sources of bias (Boutron et al., 2019; Felson, 1992; Sterne et al., 2001). The most (Bouyssou & Marchant, 2016; Perianes-Rodriguez et al., 2016; Zhou & Leydesdorff, 2010) established form is publication bias (Lortie et al., 2007; Møller & Jennions, 2001; Thornton & Lee, 2000). This form of bias is detected through funnel-plots or through weighed linear models and focusses on detecting small study effects (Egger et al., 1997; Sterne et al., 2001). This includes the absence of studies that have smaller sample sizes, possibly due to the difficulties associated with the peer-review and publishing of such studies, as well as the inclusion of studies with small sample sizes and very low variance. There is, however, another related source of bias that has received little to no recognition thus far—bias from the over-representation of studies from specific authors (Ausloos, 2013); hereafter, author bias.

Author bias, when not accounted for, has the inherent ability to skew the overview and interpretation of the literature in several ways. In the first, a specific view may be held by a particular group of authors who publish at a much higher frequency than other scientists in their field (Lortie et al., 2007), resulting in many publications in support of a view from a narrow pool of authors. This may create the illusion that opinions reported in their papers represent a majority consensus even when few independent studies support their claims. Secondly, views derived from primary findings based on a novel and ‘in-house’ method may not be fully reproducible if no independent studies exist where other authors repeated and confirmed the validity of such methods. This can further be confounded by the fact that negative or disconfirming results are often published at a delay (Boutron et al., 2019) or in less prominent journals (Leimu & Koricheva, 2004). Lastly, a majority of studies may have been conducted in a specific country, region (Collyer, 2018), and context (Fohringer et al., 2022) which—in cases where study populations may vary significantly between regions—may result in interpretations and generalisations that aren’t universal. This makes the scientometric analysis of studies by author, year, and location, critical in providing an appraisal of the literature.

At present, the most common methods used to access author contributions are the use of fractional citation counts (Bouyssou & Marchant, 2016; Perianes-Rodriguez et al., 2016; Zhou & Leydesdorff, 2010). This method emerged as a practical approach in response to the various complexities of assessing the contribution levels within scholarly works attributed to specific authors. While this method has proven useful in illuminating prominent contributors in differing fields (Bedru et al., 2023; Small & Garfield, 1985) and in citation network analyses (Perianes-Rodriguez et al., 2016), no clear link has been made between scores for individual authors from fractional counting and bias introduced in reviews from contribution levels. Furthermore, many of the methods that have been described still lack available software that implements the method (Bedru et al., 2023), or are available with very limited functionality (Keirstead, 2016; Kozłowski, 2019). To address the current need to quantitate author contribution levels as a measure of bias, and perform scientometric checks on publication year and location, ABCal (version 1.0.2) was created to compute author bias and plot scientometric aspects of studies included in systematic reviews and meta-analyses. In this paper, a full description is given of how the author bias metric is computed along with examples of how ABCal can be used to evaluate potential sources of bias using real data from two datasets from a recent systematic review (Le Clercq et al., 2023a, 2023b, 2023c).

## Methods

### Author bias metric

To assess relevant attributes of included studies, ABCal includes a new metric referred to as “author bias” to provide a quantitative measure for the possible effect of overrepresented authors introducing bias to the overall interpretation of the literature. This measure is derived in several steps. First, the full list of authors ( $L_{All}$ ) for all included studies is used to determine the total number of times,  $n_{Author}$ , the name of a specific author occurs when iterating through each position from  $i = 1$  to  $\#L_{All}$  in the list (Eq. 1).

$$n_{Author} = \sum_{i=1}^{n=\#L_{All}} f_i(Author) \quad (1)$$

This value is then divided by the total number of authors in the list,  $\#L_{All}$ , for which provides the individual author bias,  $AB_{Individual}$ , as the proportion of total authorship contributions that belong to individuals (Eq. 2).

$$AB_{Ind} = n_{Author} / \#L_{All} \quad (2)$$

Next the bias derived from authorship is calculated per study ( $AB_{Study}$ ) by adding the individual bias values,  $AB_{Ind}$ , for each author in the author list for a specific study  $L_{Paper}$ , from  $i = 1$  to the  $\#L_{Paper}$ , per Eq. 3.

$$AB_{Study} = \sum_{i=1}^{n=\#L_{Paper}} AB_{Ind}(i) \quad (3)$$

To facilitate the interpretation of these values, the final steps are calibration (Eq. 4), by dividing total bias per study,  $AB_{Study}$ , by the number of authors per paper,  $\#L_{Paper}$ .

$$AB_{Calibrated} = AB_{Study} / \#L_{Paper} \quad (4)$$

These are further categorised by assessing the distribution of author bias values by identifying those studies that fall in the bottom, middle, and upper third range, or percentiles (Eq. 5) of thirty-three, to assign bias status as being low, medium, or high based on the calculated quantiles.

$$p = x_{ri} + r_f * (x_{ri+1} - x_{ri}) \quad (5)$$

The percentile ( $p$ ) is interpolated by adding the value for  $x$  at position  $ri$  to the product of the position ( $r_f$ ) and the difference between the values for  $x$  at position  $ri + 1$  and  $ri$ . ABCal also provides some functionality to assess the normality of the author bias values using three approaches: a Shapiro–Wilk test (Shapiro & Wilk, 1965), a Quantile–Quantile (QQ) plot (Wilk & Gnanadesikan, 1968), and a histogram of distributions. The plotting sub-menu also provides an option for plotting the distribution of z-score transformed bias values which is useful in comparing distributions for different meta-analysis datasets.

The performance of the author bias metric was assessed for both validity and reliability (Cohen et al., 2017; Hammersley, 1987). Validity, in this context, refers to the ability of the metric to accurately measure the intended attribute. This was verified by comparing those studies for which a higher author bias value was computed to whether the authors

listed on the paper ranked within the top 10 contributing authors for the field. Agreement was measured in *R* 4.0.6 (R Core Team, 2020) using Cohen's kappa (Cohen, 1960) with the *vcd* 1.4–11 package (Meyer et al., 2023). Reliability or repeatability of the measure was assessed by comparing the paper level calibrated author bias levels coded as low (1), medium (2), or high (3), for two different datasets (Le Clercq et al., 2023a, 2023b, 2023c). This was done by assessing their independent validity as well as internal consistency between the distributions using Cronbach's alpha (Cronbach, 1951) with the *ltm* 1.2–0 package (Rizopoulos, 2006).

## Implementation

ABCal, version 1.0.2 (Le Clercq, 2023), was scripted in the Spyder 5 IDE using the PYTHON 3 language (Python Team, 2021) and should be compatible with all versions upward of version 3.6. The list of packages that form part of the dependencies is provided on the GitHub repository and within the README file, along with detailed instructions for the download and installation. Dependencies include the use of several core PYTHON based libraries such as *NumPy* 1.20.1 and *pandas* 1.2.4 to handle input files and mould data structures for analyses (Harris et al., 2020; McKinney, 2010). Other dependencies include packages for statistical analyses, such as *SciPy* 1.6.2 (Virtanen et al., 2020) and *statsmodels* 0.12.2 (Seabold & Perktold, 2010), and packages for graphical plotting, such as *matplotlib* 3.3.4 (Barrett et al., 2005) and *GeoPy* 2.3.0 (Lopez Gonzalez-Nieto et al., 2020). ABCal further uses the plotting functionality implemented in *folium* 0.14.0 with selected functions from the IO tools, for input and output, as well as the PYTHON Image Library (*PIL* 10.0.0). Menu options (detailed in Section "Usage") provide the utilities to calculate author bias, test the distributions of author bias values, and generate several scientometric plots. Scientometric plotting options include the ability to plot publications by the top contributing authors (to identify and visualise the extent to which top authors may skew overall interpretation), by year, and by location.

## Input and output file formats

All input files used by ABCal are in the standard comma separated value (CSV) format. For most functions the first column of these files should contain the heading "Paper" with the studies listed by name in the column e.g., "Le Clercq et al. (2023a, b, c)". To calculate the author bias, the CSV file should contain columns for each author, first to last, labelled with appropriate headings such as "Author1" etc. These columns should contain the last name and initials of each author associated with a specific paper e.g., "Le Clercq, L.S". The function to compute author bias moves through several steps to perform each intermediate step to derive the values and provides intermediate output files with relevant measures for later steps or scientometric plotting, detailed under the usage section. These files provide the option to specify unique output file names and are stored as CSV files within the current working directory.

For most of the plotting options, either the CSV files generated from author bias computation or CSV files containing additional study attributes for plotting are used. An example of such a file to plot the distribution of publications by year, includes a file containing two columns with the headers "Paper" and "Year", which should contain the study name as well as the year of publication. Another example is for the plotting of studies by location, where a CSV is required containing two columns with the headers "Paper" and "Location".

For this file, the full name of the country in which the study was conducted is required; if more than one location was included these should be listed on separate lines with the study name and second or third location. The output generated from plotting is saved with a standard name for the type of plot in the portable network graphic (PNG) format, with the exception of the location plot which is also stored in the interactive HTML format.

## Usage

To illustrate the usage of ABCal, two datasets (Le Clercq et al., 2023a, 2023b, 2023c) generated as part of a recent systematic review on biomarkers for age in animals (Le Clercq et al. 2023c), comprising age models from included studies on the use of methylation ( $N=41$  studies, 60 models) and telomeres ( $N=67$  studies, 99 models) respectively, will be used. For each dataset, three input files were generated: one with the paper name and list of authors, a second with the paper name and publication date, and a third with the paper name and study location.

The first file was used to compute the author bias (example A). Once ABCal is initiated, the first option (a) is to calculate the total author bias per paper. Selecting this option initiated the function to perform the needed steps to do the calculation. A prompt appeared to specify the name of the file containing the author lists e.g., “Auth\_Meth.csv”. The first step generated a list of all authors along with the total counts of times the specific author appeared in an author list. This data was exported as the first file for output from the function and was saved as a CSV containing the number of publications per author. Next, the individual author bias was computed by determining the fraction of total authorship contributions per author. These values were stored in the second output and contained the author names and their associated individual bias. The final step computed the total author bias per paper by adding the individual bias value of each author associated with the author list for a paper. This data was saved as the third output and contained the data used to compute total bias per paper. As an additional step, and to assist in the interpretation of values, the second menu option (b), which takes the final output file from the first option, was used to calibrate the bias value by dividing the total bias by the number of authors per paper. The newly calibrated values were exported as the fourth output file. Furthermore, the third menu option (c) was used for testing the distributions for normality and the fourth menu option (d) was used to get the upper, middle, and lower third quantiles of the author bias distributions.

Hereafter, the author bias values and their respective levels were incorporated into two meta-analyses as part of a review (Le Clercq et al. 2023c). The meta-analysis was done in *RStudio 1.4.1106* (RStudio Team, 2021), running *R 4.0.5* (R Core Team, 2020) with the package *meta 5.5-0* (Harrer et al., 2021; Schwarzer et al., 2015). A meta-regression was done between the random effects model and author bias as a predictor of heterogeneity for a functional test of validity. The results were visualised using a bubble plot implemented in the *metafor 3.8-0* package (Viechtbauer, 2010) with grouping based on the three quantiles. Furthermore, potential publication bias as measured by funnel plot asymmetry was also assessed using the Egger’s test (Egger et al., 1997) as implemented in *metafor 3.8-0*, also plotting the relationship between the standardized measured effect and the inverse of the standard error.

Scientometric plotting capabilities of ABCal, accessed via a submenu when selecting option e, were illustrated in example B for methylation studies and example C for telomere studies. For the first option (a) from the submenu, the second file with two columns for

‘Paper’ and ‘Year’ was used to plot the total number of publications per year. The first file given as output from the author bias computation steps, containing the authorship counts, was used for the second option (b) to plot the number of contributions from the top ten contributing authors. Lastly, the third file containing two columns for ‘Paper’ and ‘Location’ was used to plot a choropleth map of the geographical distribution for study locations using the third option (c) from the submenu.

## Results

### Example A: author bias computation

Author bias values per paper for methylation studies ranged from 0.0024 to 0.0302 with a mean of 0.0124 (Table 1). The histogram plot of distributions (Fig. 1A) showed many values ( $N=19$ ) were well below the mean, skewing the distribution left, with a moderate number of studies ( $N=10$ ) falling in and around the mean and few studies ( $N=12$ ) having higher values. The position for setting the first (Q1) and third (Q3) quartiles were 0.004 and 0.018 respectively, with 14 studies classified as low, 13 as medium, and 14 as high. The box plot of z-score transformed values (Fig. 1B), to express the bias values in terms of standard deviations from the median, showed that the median was low (approximately  $-0.094$ ) with most studies (95%) being evenly distributed around the median. A few studies had higher values; however, they still fell within two standard deviations of the median and no clear outliers were detected. The overall distributions were found to not be normally distributed (Table 1;  $P<0.01$ ). Tests for validity by means of Cohen’s kappa showed high

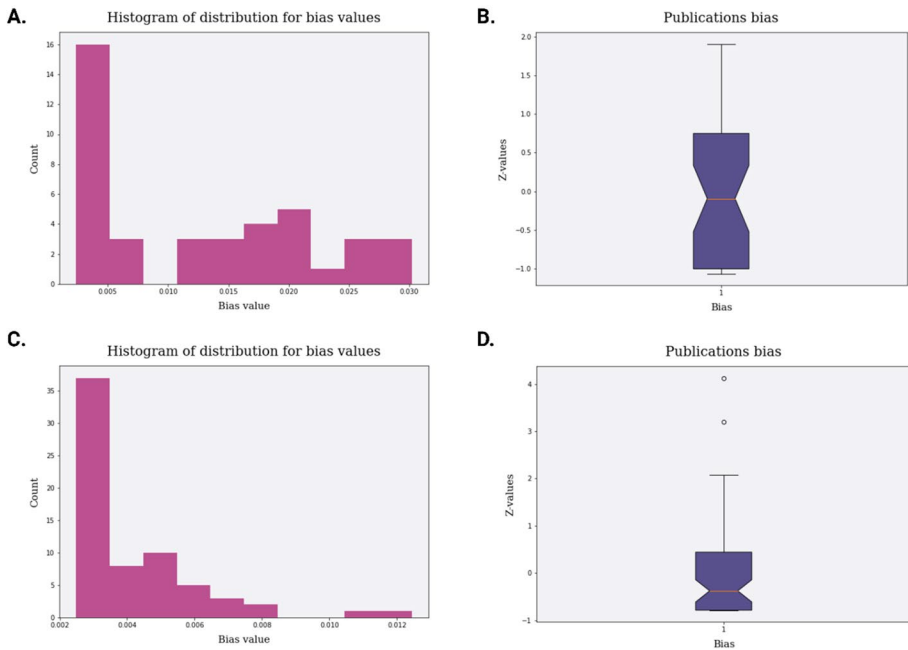
**Table 1** Summary of characteristics of calibrated author bias values

	Mean	SE	Min	Max	Q1	Q3	Shapiro–Wilk
<b>Methylation</b>							
Author bias	0.014	0.001	0.002	0.030	0.004	0.018	0.869 ( $P<0.01$ )
Z-score	$-1.08 \times 10^{16}$	0.16	$-1.07$	1.90	$-1.00$	0.75	
(1) Low	14						
(2) Medium	13						
(3) High	14						
<b>Telomeres</b>							
Author bias	0.004	0.0001	0.002	0.012	0.003	0.004	0.780 ( $P<0.01$ )
Z-score	$-2.10 \times 10^{16}$	0.12	$-0.79$	4.13	$-0.79$	0.44	
(1) Low	23						
(2) Medium	21						
(3) High	23						

Values are reported for both the calibrated author bias (AB) values as well as the Z-Score transformed values. The mean, standard error of the mean (SE), minimum (Min), maximum (Max) and positions of the cut-off points for the lower (Q1) and upper (Q3) third percentiles are given. The results for the normality test, tested using the Shapiro–Wilk test, are also given along with the significance. The calibrated author bias for both datasets was not normally distributed ( $P<0.01$ ). The number of studies identified by level as having either low, medium, or high risk of bias are also indicated



# Plots for Distributions of Author Bias values



**Fig. 1** Plots for the distributions of author bias values. **A** Histogram for the calibrated author bias values for papers in the methylation dataset indicating many studies ( $N=16$ ) with low values, skewing the distribution left, with a moderate number of studies ( $N=4-6$ ) around the median and a similar number of studies ( $N=3-6$ ) with high values. **B** Box plot for the Z-score transformed author bias values for papers in the methylation dataset indicating most studies are evenly distributed around the median (orange line) with few studies more than one standard deviation from the median, and a small number of studies in the upper range without appearing as outliers. **C** Histogram for the calibrated author bias values for papers in the telomere dataset indicating many studies ( $N=25$ ) with low values, skewing the distribution left, with a moderate number of studies ( $N=5-10$ ) around the median and two studies with higher values. **D** Box plot for the Z-score transformed author bias values for papers in the telomere dataset indicating most studies were slightly above the median (orange line) with most studies within one standard deviation from the median, and a small number of studies in the upper range and two outliers. (image created in BioRender.com)

levels of agreement (Table 2;  $\kappa=0.94$ ,  $P<0.01$ ) between studies ranked as having medium to high risk of bias as compared to the list of top contributing authors.

Author bias values for telomere studies ranged from 0.0024 to 0.0124 with a mean of 0.0040 (Table 1). The position for setting the first (Q1) and third (Q3) quartiles were 0.003 and 0.004 respectively, with 23 studies classified as low, 21 as medium, and 23 as high. The values followed a similar pattern to that observed for methylation studies (Fig. 1C), with a large number of studies ( $N=48$ ) falling below the mean. Most of the remaining studies followed a near bell shape around the mean, with several ( $N=9$ ) having higher values. The box plot (Fig. 1D) showed a slightly higher number of studies fell above the median, while most studies were still within one standard deviation of the median. Several studies had values between one and two standard deviations of the median, with two studies that were more than two standard deviations from the median and thus detected as outliers. Once more, the distributions were found to not be normally distributed (Table 1;  $P<0.01$ ). Tests for validity showed a moderate, yet significant, level of agreement (Table 2;  $\kappa=0.52$ ,

**Table 2** Results from tests for validity, by mean of Cohen's kappa ( $\kappa$ ), as well as reliability, by means of Cronbach's alpha ( $\alpha$ ), for the Author Bias metric

	Value	SE	z-value	Confidence
Validity				
<i>Methylation</i>	$\kappa$ : 0.94	0.05	17.36	$P < 0.01$
<i>Telomeres</i>	$\kappa$ : 0.52	0.13	3.89	$P < 0.01$
Between	$\kappa$ : 0.73	0.09		Good
Reliability				
Between	$\alpha$ : 0.84	–	–	CI 0.74–0.91

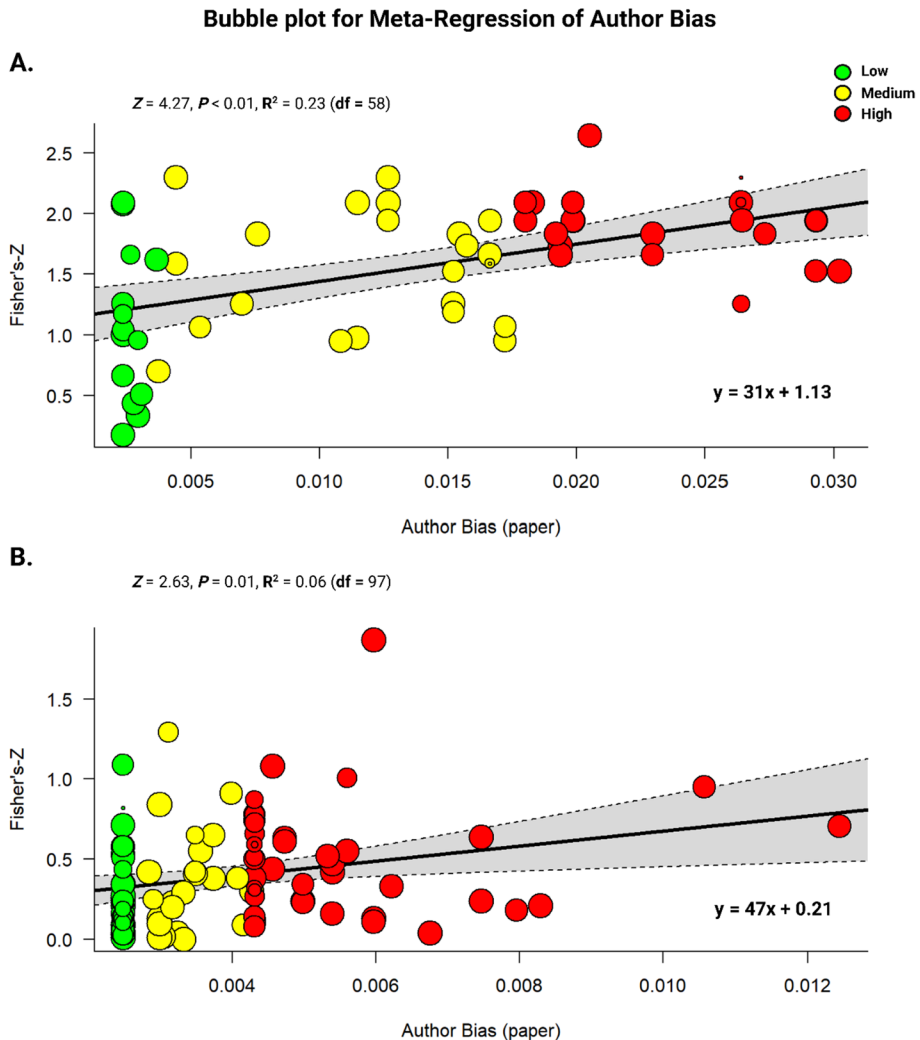
Validity was calculated for both the methylation and telomere dataset by assessing the agreement between bias rank and authors listed as the top ten contributors. The overall validity was determined by taking the average for individual results. Agreement for ranking between datasets was used to assess between study reliability of the metric. The confidence of the calculated values was assessed by either probability ( $P < 0.01$ ) of a z-test or the 98% confidence interval (CI)

$P < 0.01$ ) and an average validity between datasets of 0.73. The reliability tests between datasets also showed a significant (Table 2;  $\alpha = 0.84$ , CI 0.74–0.91) level of reproducibility for levelled classification of studies.

A meta-regression between observed effect measures and author bias values revealed that author bias values were able to account for a significant portion of the observed heterogeneity between studies included in meta-analyses. These relationships are illustrated as bubble plots in Fig. 2. For methylation studies, a strong relationship was observed ( $P < 0.01$ ) with author bias values accounting for approximately 23% ( $R^2 = 0.23$ ) of the study heterogeneity. For telomere studies, a slightly weaker relationship was observed ( $P < 0.02$ ) with author bias values accounting for 6% ( $R^2 = 0.06$ ) of the study heterogeneity. In both instances, higher levels of author bias were associated with higher effect measures while lower author bias was evident for studies with lower effect measures. Tests for publication bias (Fig. S1) detected significant funnel plot asymmetry ( $P < 0.05$ ) indicative of possible publication bias. Statistical methods to address publication bias, such as “trim-and-fill” or linear modelling of a fixed-effect model with factorisation, did not significantly alter the overall interpretations from the comparisons (data not shown).

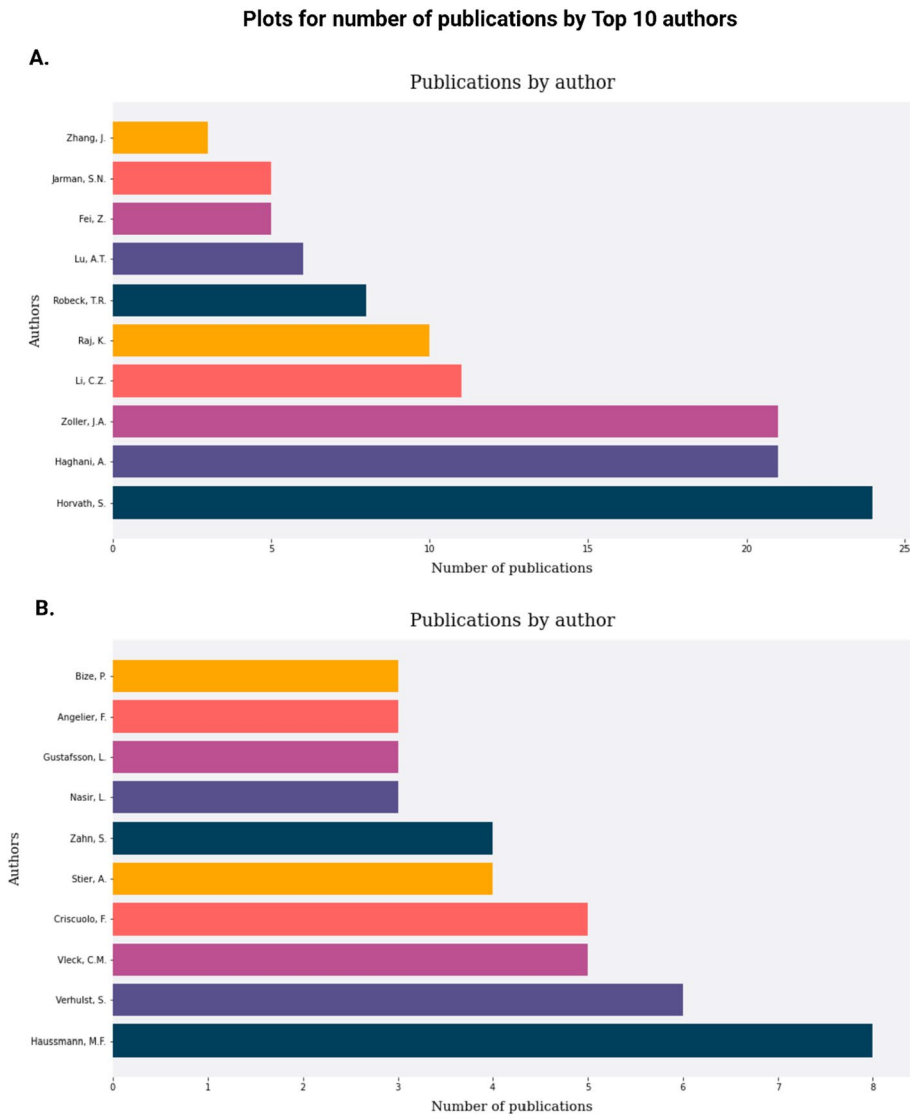
## Example B: a meta-analysis of methylation studies

Scientometric assessment of studies included in the methylation dataset was done by plotting three attributes: top contributing authors, publications by year, and publications by location. For publications per author, the number of publications contributed by the top contributing authors (specified as ten) ranged from 3 contributions to a total of 24 contributions (Fig. 3A). Five authors, including Zhang, contributed 3 papers each, respectively. The top three contributing authors, identified as Horvath, Haghani, and Zoller, contributed to approximately half (21–24 out of 41) of the included studies. The bar plot for publication by year (Fig. 4A) revealed the first studies were published in 2014 (Polanowski et al., 2014) with an annual increase leading to 15 publications in 2021 (Bors et al., 2021; Mayne et al., 2021; Robeck et al., 2021; Wilkinson et al., 2021) and several recent studies (Horvath et al., 2022a, 2022b; Horvath et al., 2022a, 2022b; Robeck et al., 2023). The choropleth map (Fig. 4B), showing study locations, showed



**Fig. 2** Bubble plots for the meta-regression of author bias values as a predictor of heterogeneity in the meta-analyses. The Fisher's-Z values (y-axis) were plotted against the calibrated author bias values per paper (x-axis) and studies colour coded according to the quantiles within which they fell and were classified: low (green), medium (yellow), or high (red). The linear equations for the regressions are indicated in the bottom right of each plot. **A** Meta-regression performed on the methylation dataset found a significant association between effect measures and author bias values ( $P < 0.01$ ), accounting for 23 percent of the observed heterogeneity. **B** Meta-regression for the telomere dataset found a moderate association ( $P < 0.02$ ), accounting for only 6 percent of the heterogeneity. (image created in BioRender.com)

the number of studies per country ranged from one study (green) to more than twenty studies (red). Several countries, shown in white, were completely data deficient. The overall distribution showed that most studies emanated from the Northern hemisphere, principally from North America ( $N > 20$ ) and Europe, with Australia ( $N > 5$ ) representing the country with the most publications in the global South.

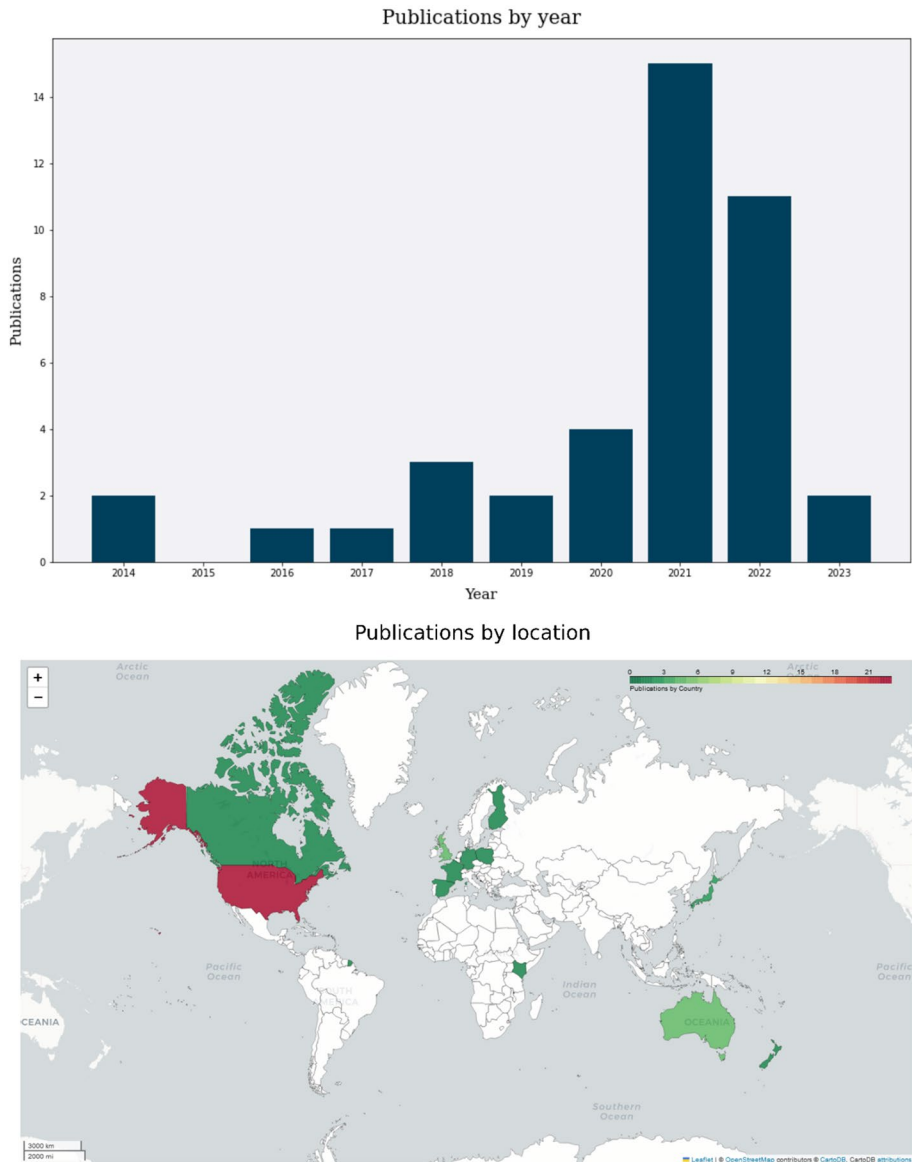


**Fig. 3** Scientometric plots for the number of publications attributed to the top 10 contributing authors. **A** For methylation studies, the number of publications attributed to top contributing authors ranged from 4 contributions to as many as 24 contributions: with the top 3 contributing to a half (50%) of the included studies. **B** For telomere studies, the number of publications attributed to the top contributing authors ranged from 3 to 8 contributions: here, the top 3 contributing authors made up approximately a tenth (10%) of the total studies. (image created in BioRender.com)

### Example C: a meta-analysis of telomere studies

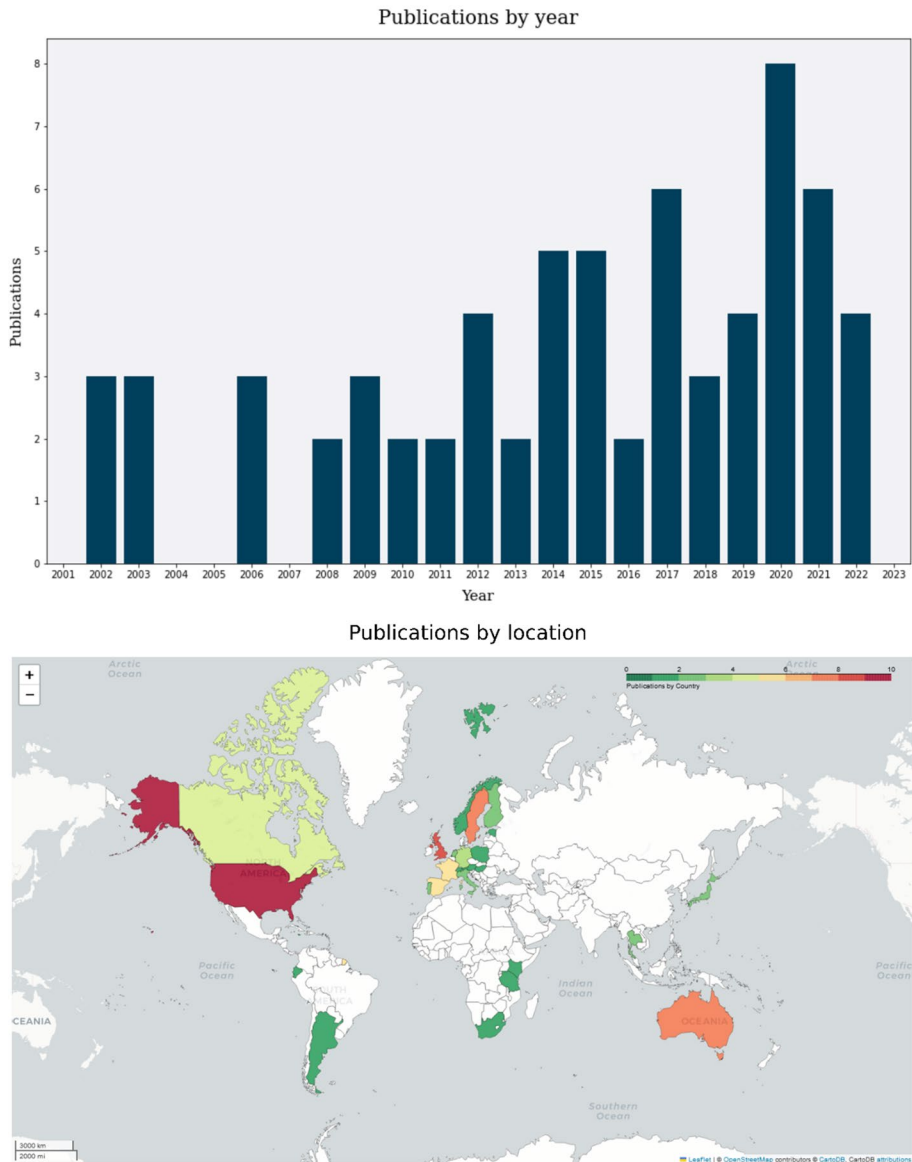
The same scientometric plots were also generated for the telomere dataset. For publications per author, contributions by the top ten authors ranged from 3 contributions to a total of 8 contributions (Fig. 3B). Several authors contributed 3 papers while the top three authors,

# Scientometric plots for Methylation studies



**Fig. 4** Scientometric plots for methylation studies, generated with ABCal. **A** Bar plot for publication by year, indicating the first studies published in 2014 with an annual increase leading to 15 publications in 2021. **B** Choropleth map showing study locations. The density gradient plots the number of studies per country ranging from one study (green) to more than twenty studies (red); countries in white are data deficient. The overall distribution shows most studies are from the Northern hemisphere, principally from North America and Europe, as well as Australia. (image created in BioRender.com). (Color figure online)

# Scientometric plots for Telomere studies



**Fig. 5** Scientometric plots for telomere studies, generated with ABCal. **A** Bar plot for publications by year, showing the first included studies were published in 2002 with constant publication of 2–3 studies per year and increasing from 2012 with several spikes in 2017, 2020, and 2021 to between 6–8 publications. **B** Choropleth map showing study locations. The density gradient plots the number of studies per country ranging from one study (green) to ten studies (red); countries in white are data deficient. The distribution shows most studies are from the Northern hemisphere, principally from North America, as well as Australia. (image created in BioRender.com)

identified as Haussmann, Verhulst, Vleck, and Criscuolo, each contributed between 5 and 8 studies. This only accounted for about 10% of the included studies (5–8 out of 68). Publications by year, given as a bar plot (Fig. 5A), indicated the first included studies were published circa 2002 (Brümmendorf et al., 2002; Haussmann & Vleck, 2002) with frequent subsequent publications, around 2–3 studies annually, and an increase seen after 2012 (Fick et al., 2012; Plot et al., 2012) with several spikes in 2017 (Cerchiara et al., 2017; Kirby et al., 2017; Ujvari et al., 2017), 2020 (Bauch et al., 2020; Burraco et al., 2020; Cherdasukjai et al., 2020), and 2021 (Molbert et al., 2021; Vernasco et al., 2021). The maximum number for spikes ranged between 6 to 8 publications. Choropleth mapping of study locations as publications per country (Fig. 5B) ranged from one study (green) to ten studies (red); data deficient countries are indicated in white. The distribution showed most studies originated from the Northern hemisphere, principally from North America ( $N > 9$ ). The most represented country from the Southern hemisphere was Australia ( $N > 6$ ).

## Discussion

This original paper presents the methods and results of a research study using a newly developed software tool called ABCal. The tool is implemented in Python and is designed to analyse scientometric data from various studies when conducting systematic reviews and meta-analyses. The primary focus of the study was to compute and assess author bias in the literature, providing a quantitative measure for the possible effect of overrepresented authors introducing bias to the overall interpretation of the literature. The computed author bias values provide a quantitative measure of author influence on the interpretation of the studies. Furthermore, scientometric plots provided valuable insights into the trends and distribution of publications over time and geographic locations.

The distribution of author bias values, as shown in histograms and box plots, showed marginal differences in the raw values but was conserved between the two datasets when using z-score transformed values. This highlighted similar spreads for the distribution with low to medium bias for most studies and a smaller number of studies exhibiting higher bias levels. As such, author bias values were able to identify, in a quantitative manner, the overrepresentation of some authors in both meta-analytic datasets. When combined with scientometric plots of author contributions, it became clear that both datasets contain a significant number of authors who have contributed a larger number of studies than others. This makes the proposed author bias metric useful in addressing the account for such bias when doing reviews (Felson, 1992; Knobloch et al., 2011).

The author bias metric is organically related to fractional citation counting (Zhou & Leydesdorff, 2010), an emerging ‘golden standard’ when comparing author contribution levels. This is due to similarities between calculations used in the initial steps that count the total number of times an author appears in the list of authors, which is divided by the total number of authors in the list (Bouyssou & Marchant, 2016). Considering, however, that this is only done within the context of studies included in a meta-analysis instead of the full reference list—in this instance the new metric represents a special use case of fractional counting. Subsequent steps sum the fractional count for individual authors per included study and divides the total by the number of authors per paper to derive a paper-level value for author bias. Furthermore, rather than relying on raw values, ABCal provides the option to convert between raw values and z-scores as well as three levels of interpretation: low, medium, and high. This facilitates cross-discipline

use of the calibrated author bias metric as several cultural differences may exist between fields in terms of publication and citation behaviour (Bornmann & Daniel, 2008; Zhou & Leydesdorff, 2010). At present ABCal, and the novel author bias metric, also provides enhanced utility in comparison to existing options (Bedru et al., 2023; Keirstead, 2016; Kozłowski, 2019). For example, ABCal uses provided information and does not rely on the indexing of papers on a specific database (Kozłowski, 2019) or the existence of author profiles on a specific platform (Keirstead, 2016). This is particularly important for included studies from publishers that don't index their articles on all databases or when including preprints from e.g., *bioRxiv*. ABCal is also freely available to the community for implementation in other studies while several similar algorithms are not (Bedru et al., 2023).

Any new metric is, however, subject to benchmarking through tests of validity and reliability (Cohen et al., 2017; Hammersley, 1987). Validity, as measured by agreement between medium to high bias studies and the list of top authors using Cohen's kappa (Cohen, 1960), showed moderate to high agreement levels that are generally well suited given the application (Altman, 1990). Reliability, as measured by Cronbach's alpha, found that the author bias metric was able to partition studies into low, medium, and high bias with a high degree of consistency between datasets. Furthermore, the functional validity was assessed by meta-regression for which the results indicated a significant association between author bias and observed effect measures in the meta-analyses. More specifically, higher author bias values were associated with higher effect measures, while lower bias was evident in studies with lower effect measures. This also makes author bias values utile in understanding how author prominence (Cassey et al., 2004), from higher contributions to the field, may interact with reported effect sizes to account for part of the heterogeneity that exists between studies as well as publication bias (Baker et al., 2009).

The presence of publication bias, as revealed by tests of funnel plot asymmetry (Møller & Jennions, 2001), suggests the possibility of selective publication in the literature (Boutron et al., 2019). This is typically attributed to small study effects such as the exclusion of studies with smaller sample sizes, even when sample sizes are sufficient for adequate statistical power of a given test (Cohen, 1988; Faul et al., 2009; Kang, 2021). The concentration of studies in certain regions as seen by geographic mapping of study locations, however, indicates a potential research trend of fewer or missing studies from the global South, as previously suggested (Collyer, 2018), and evidence that research in ecology or animal science may not follow a truly global distribution (Martin et al., 2012). Considering researchers from lower income countries may conduct research on a smaller scale for economic reasons, it is feasible that the existing evidence of publication bias is due to the lack of studies from the Southern hemisphere in the literature.

It's important to acknowledge the limitations of the study. The analysis relies on the accuracy and completeness of the input data (Knobloch et al., 2011; O'Dea et al., 2021), and certain assumptions might have been made during the calculation of author bias. Additionally, the analysis is limited to the specific datasets related to biomarkers for age in animals, and generalization to other research fields might require further investigation. Future work can focus on expanding the application of ABCal to different research areas and datasets to validate its effectiveness and robustness across various domains. Additionally, efforts can be made to address potential limitations and explore enhancements to the tool's functionalities to meet the evolving needs of scientometric analysis in the ecology research community, particularly when conducting systematic reviews and meta-analyses.

Overall, ABCal proves to be a useful tool for scientometric analysis, offering valuable information to researchers in assessing the impact of authors and potential biases in the



literature. The software's capabilities to analyse authorship contributions and produce scientometric plots can aid researchers in gaining a deeper understanding of the research landscape and identifying key contributors and research trends.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s11192-023-04880-6>.

**Acknowledgements** All images were edited for publication in BioRender.com with publication licenses granted under the Academic plan. The author would like to thank Dr Desiré Lee Dalton, and Professors Paul Grobler and Antoinette Kotzé, for their support and advice.

**Funding** Open access funding provided by University of the Free State. Open access provided by the University of the Free State. This work is based on the research supported wholly/in part by the National Research Foundation (Grant Number: 112062), South Africa.

**Data availability** The custom Python script for ABCal version 1.0.2 is available for download for installation from source code on GitHub (<https://github.com/LSLeClercq/ABCal>), and includes example files used for testing. Data used in this paper were deposited online (<https://doi.org/10.5281/zenodo.7091053>) and recently published (Le Clercq et al., 2023a, 2023b, 2023c).

## Declarations

**Conflict of interest** The author has no competing interests to declare.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Adams, D. C. (2008). Phylogenetic meta-analysis. *Evolution*, 62(3), 567–572. <https://doi.org/10.1111/j.1558-5646.2007.00314.x>
- Altman, D. (1990). *Practical statistics for medical research* (1st ed.). London: Chapman & Hall.
- Ausloos, M. (2013). A scientometrics law about co-authors and their ranking: The co-author core. *Scientometrics*, 95(3), 895–909. <https://doi.org/10.1007/s11192-012-0936-x>
- Baker, W. L., Michael White, C., Cappelleri, J. C., Kluger, J., & Coleman, C. I. (2009). Understanding heterogeneity in meta-analysis: The role of meta-regression. *International Journal of Clinical Practice*, 63(10), 1426–1434. <https://doi.org/10.1111/J.1742-1241.2009.02168.X>
- Barrett, P., Hunter, J., Miller, J. T., Hsu, J.-C., & Greenfield, P. (2005). matplotlib—A portable python plotting package. In *Astronomical data analysis software and systems XIV ASP conference series* (Vol. 347, p. 91). Pasadena
- Bauch, C., Gatt, M. C., Granadeiro, J. P., Verhulst, S., & Catry, P. (2020). Sex-specific telomere length and dynamics in relation to age and reproductive success in Cory's shearwaters. *Molecular Ecology*, 29(7), 1344–1357. <https://doi.org/10.1111/mec.15399>
- Bedru, H. D., Zhang, C., Xie, F., Yu, S., & Hussain, I. (2023). CLARA: Citation and similarity-based author ranking. *Scientometrics*, 128(2), 1091–1117. <https://doi.org/10.1007/s11192-022-04590-5>
- Boell, S. K., & Cecez-Kecmanovic, D. (2010). Literature reviews and the hermeneutic circle. *Australian Academic and Research Libraries*, 41(2), 129–144. <https://doi.org/10.1080/00048623.2010.10721450>
- Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2010). A basic introduction to fixed-effect and random-effects models for meta-analysis. *Research Synthesis Methods*, 1(2), 97–111. <https://doi.org/10.1002/JRSM.12>

- Borenstein, M., & Higgins, J. P. T. (2013). Meta-analysis and subgroups. *Prevention Science*, 14(2), 134–143. <https://doi.org/10.1007/S11121-013-0377-7/METRICS>
- Bornmann, L., & Daniel, H. D. (2008). What do citation counts measure? A review of studies on citing behavior. *Journal of Documentation*, 64(1), 45–80. <https://doi.org/10.1108/00220410810844150/FULL/XML>
- Bors, E. K., Baker, C. S., Wade, P. R., O'Neill, K. B., Shelden, K. E. W., Thompson, M. J., Fei, Z., Jarman, S., & Horvath, S. (2021). An epigenetic clock to estimate the age of living beluga whales. *Evolutionary Applications*, 14(5), 1263–1273. <https://doi.org/10.1111/eva.13195>
- Boutron, I., Page, M., Higgins, J., Altman, D., Lundh, A., & Hróbjartsson, A. (2019). Chapter 7: Considering bias and conflicts of interest among the included studies. In J. Higgins, J. Thomas, J. Chandler, M. Cumpston, T. Li, M. Page, & V. Welch (Eds.), *Cochrane handbook for systematic reviews of interventions* (2nd ed., pp. 177–204). Chichester: Wiley.
- Bouyssou, D., & Marchant, T. (2016). Ranking authors using fractional counting of citations: An axiomatic approach. *Journal of Informetrics*, 10(1), 183–199. <https://doi.org/10.1016/J.JOI.2015.12.006>
- Brümmendorf, T. H., Mak, J., Sabo, K. M., Baerlocher, G. M., Dietz, K., Abkowitz, J. L., & Lansdorp, P. M. (2002). Longitudinal studies of telomere length in feline blood cells: Implications for hematopoietic stem cell turnover in vivo. *Experimental Hematology*, 30(10), 1147–1152. [https://doi.org/10.1016/S0301-472X\(02\)00888-3](https://doi.org/10.1016/S0301-472X(02)00888-3)
- Burraco, P., Comas, M., Reguera, S., Zamora-Camacho, F. J., & Moreno-Rueda, G. (2020). Telomere length mirrors age structure along a 2200-m altitudinal gradient in a Mediterranean lizard. *Comparative Biochemistry and Physiology Part a: Molecular & Integrative Physiology*, 247, 110741. <https://doi.org/10.1016/j.cbpa.2020.110741>
- Cassey, P., Ewen, J. G., Blackburn, T. M., & Møller, A. P. (2004). A survey of publication bias within evolutionary ecology. *Proceedings of the Royal Society B Biological Sciences*. <https://doi.org/10.1098/rsbl.2004.0218>
- Cerchiara, J. A., Risques, R. A., Prunkard, D., Smith, J. R., Kane, O. J., & Boersma, P. D. (2017). Magellanic penguin telomeres do not shorten with age with increased reproductive effort, investment, and basal corticosterone. *Ecology and Evolution*, 7(15), 5682–5691. <https://doi.org/10.1002/ECE3.3128>
- Chamberlain, S. A., Hovick, S. M., Dibble, C. J., Rasmussen, N. L., Van Allen, B. G., Maitner, B. S., Ahern, J. R., Bell-Dereske, L. P., Roy, C. L., Meza-Lopez, M., & Carrillo, J. (2012). Does phylogeny matter? Assessing the impact of phylogenetic information in ecological meta-analysis. *Ecology Letters*. <https://doi.org/10.1111/j.1461-0248.2012.01776.x>
- Cherdsukjai, P., Buddhachat, K., Brown, J., Kaewkool, M., Poommouang, A., Kaewmong, P., Kittiwattana-wong, K., & Nganvongpaint, K. (2020). Age relationships with telomere length, body weight and body length in wild dugong (Dugong dugon). *PeerJ*, 8, e10319. <https://doi.org/10.7717/peerj.10319>
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1), 37–46. <https://doi.org/10.1177/001316446002000104>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.
- Cohen, L., Manion, L., & Morrison, K. (2017). Validity and reliability. In *Research methods in education* (Vol. 44, pp. 245–284). Eighth edition. New York: Routledge
- Collyer, F. M. (2018). Global patterns in the publishing of academic knowledge: Global North, global South. *Current Sociology*, 66(1), 56–73. <https://doi.org/10.1177/0011392116680020>
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3), 297–334. <https://doi.org/10.1007/BF02310555>
- Dickersin, K., Scherer, R., & Lefebvre, C. (1994). Systematic reviews: Identifying relevant studies for systematic reviews. *BMJ*, 309(6964), 1286. <https://doi.org/10.1136/bmj.309.6964.1286>
- Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *BMJ*, 315(7109), 629–634. <https://doi.org/10.1136/BMJ.315.7109.629>
- Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G\*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, 41(4), 1149–1160. <https://doi.org/10.3758/BRM.41.4.1149>
- Felson, D. T. (1992). Bias in meta-analytic research. *Journal of Clinical Epidemiology*, 45(8), 885–892. [https://doi.org/10.1016/0895-4356\(92\)90072-U](https://doi.org/10.1016/0895-4356(92)90072-U)
- Fick, L. J., Fick, G. H., Li, Z., Cao, E., Bao, B., Heffelfinger, D., Parker, H.G., Ostrander, E.A., & Riabowol, K. (2012). Telomere length correlates with life span of dog breeds. *Cell Reports*, 2(6), 1530–1536. <https://doi.org/10.1016/J.CELREP.2012.11.021>
- Fohringer, C., Hoelzl, F., Allen, A. M., Cayol, C., Ericsson, G., Spong, G., Smith, S., & Singh, N. J. (2022). Large mammal telomere length variation across ecoregions. *BMC Ecology and Evolution*, 22(1), 1–10. <https://doi.org/10.1186/S12862-022-02050-5/TABLES/1>

- Hammersley, M. (1987). Some notes on the terms 'validity' and 'reliability.' *British Educational Research Journal*, 13(1), 73–82. <https://doi.org/10.1080/0141192870130107>
- Harrer, M., Cuijpers, P., Furukawa, T. A., & Ebert, D. D. (2021). *Doing meta-analysis with R: A hands-on guide* (1st ed.). Boca Raton: Chapman & Hall/CRC Press.
- Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., & Kern, R. (2020). Array programming with NumPy. *Nature*, 585(7825), 357–362. <https://doi.org/10.1038/s41586-020-2649-2>
- Hausmann, M. F., & Vleck, C. M. (2002). Telomere length provides a new technique for aging animals. *Oecologia*, 130(3), 325–328. <https://doi.org/10.1007/s00442-001-0827-y>
- Higgins, J. P. T. (2008). Commentary: Heterogeneity in meta-analysis should be expected and appropriately quantified. *International Journal of Epidemiology*, 37(5), 1158–1160. <https://doi.org/10.1093/ije/dyn204>
- Higgins, J. P. T., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine*, 21(11), 1539–1558. <https://doi.org/10.1002/sim.1186>
- Hono, G., Leclercq, V., Geerinck, A., Thomas, T., Veronese, N., Charles, A., Rabenda, V., Beaudart, C., Cooper, C., Reginster, J. Y., & Bruyère, O. (2019). Safety of topical non-steroidal anti-inflammatory drugs in osteoarthritis: Outcomes of a systematic review and meta-analysis. *Drugs and Aging*, 36(1), 45–64. <https://doi.org/10.1007/s40266-019-00661-0>
- Horvath, S., Haghani, A., Peng, S., Hales, E. N., Zoller, J. A., Raj, K., Larison, B., Robeck, T. R., Petersen, J. L., Bellone, R. R., & Finno, C. J. (2022a). DNA methylation aging and transcriptomic studies in horses. *Nature Communications*, 13(1), 1–13. <https://doi.org/10.1038/s41467-021-27754-y>
- Horvath, S., Haghani, A., Zoller, J. A., Raj, K., Sinha, I., Robeck, T. R., Black, P., Couzens, A., Lau, C., Manoyan, M., & Ruiz, Y. A. (2022b). Epigenetic clock and methylation studies in marsupials: Opossums, Tasmanian devils, kangaroos, and wallabies. *GeroScience*, 44(3), 1825–1845. <https://doi.org/10.1007/s11357-022-00569-5>
- Kang, H. (2021). Sample size determination and power analysis using the G\*Power software. *Journal of Educational Evaluation for Health Professions*. <https://doi.org/10.3352/JEEHP.2021.18.17>. National Health Personnel Licensing Examination Board of the Republic of Korea.
- Keirstead, J. (2016). scholar: analyse citation data from Google Scholar. R package. <https://github.com/jkeirstead/scholar>
- Kirby, R., Alldredge, M. W., & Pauli, J. N. (2017). Environmental, not individual, factors drive markers of biological aging in black bears. *Evolutionary Ecology*, 31(4), 571–584. <https://doi.org/10.1007/s10682-017-9885-4>
- Knobloch, K., Yoon, U., & Vogt, P. M. (2011). Preferred reporting items for systematic reviews and meta-analyses (PRISMA) statement and publication bias. *Journal of Cranio-Maxillofacial Surgery*, 39(2), 91–92. <https://doi.org/10.1016/j.jcms.2010.11.001>
- Kozłowski, L. P. (2019). fCite: a fractional citation tool to quantify an individual's scientific research output. *bioRxiv*. <https://doi.org/10.1101/771485>
- Kraus, S., Breier, M., Lim, W. M., Dabić, M., Kumar, S., Kanbach, D., & Ferreira, J. J. (2022). Literature reviews as independent studies: Guidelines for academic practice. *Review of Managerial Science*, 16(8), 2577–2595. <https://doi.org/10.1007/s11846-022-00588-8>
- Lajeunesse, M. J. (2009). Meta-analysis and the comparative phylogenetic method. *American Naturalist*, 174(3), 369–381. <https://doi.org/10.1086/603628>
- Le Clercq, L. S. (2023). ABCal: Author bias computation and scientometric plotting. GitHub. <https://github.com/LSLeClercq/ABCal>
- Le Clercq, L., Bazzi, G., Cecere, J. G., Gianfranceschi, L., Grobler, J. P., Kotzé, A., Rubolini, D., Liedvogel, M., & Dalton, D. L. (2023a). Time trees and clock genes: A systematic review and comparative analysis of contemporary avian migration genetics. *Biological Reviews*, 98(4), 1051–1080. <https://doi.org/10.1111/brv.12943>
- Le Clercq, L.-S., Grobler, J. P., Kotzé, A., & Dalton, D. L. (2023b). Dataset generated in a systematic review and meta-analysis of biological clocks as age estimation markers in animal ecology. *Data in Brief*, 51, 109615. <https://doi.org/10.1016/j.dib.2023.109615>
- Le Clercq, L., Kotzé, A., Grobler, J. P., & Dalton, D. L. (2023c). Biological clocks as age estimation markers in animals: A systematic review and meta-analysis. *Biological Reviews*, 98(6), 1972–2011. <https://doi.org/10.1111/brv.12992>
- Le Clercq, C. M. P., van Ingen, G., Ruytjens, L., & van der Schroeff, M. P. (2016). Music-induced hearing loss in children, adolescents, and young adults. *Otology & Neurotology*, 37(9), 1208–1216. <https://doi.org/10.1097/MAO.0000000000001163>

- Leimu, R., & Koricheva, J. (2004). Cumulative meta-analysis: A new tool for detection of temporal trends and publication bias in ecology. *Proceedings of the Royal Society B: Biological Sciences*, 271(1551), 1961–1966. <https://doi.org/10.1098/rspb.2004.2828>
- Lopez Gonzalez-Nieto, P., Gomez Flechoso, M., Arribas Mocoroa, M. E., Muñoz Martin, A., Garcia Lorenzo, M. L., Cabrera Gomez, G., Gomez, J.A., Fraile, A.C., Dagan, J.O., Palomares, R.M., & Lahoz-Beltra, R. (2020). Design and development of a virtual laboratory in PYTHON for the teaching of data analysis and mathematics in geology: GeoPy. In *INTED2020 proceedings* (pp. 2236–2242). Valencia, Spain. <https://doi.org/10.21125/inted.2020.0687>
- Lortie, C. J., Aarssen, L. W., Budden, A. E., Koricheva, J. K., Leimu, R., & Tregenza, T. (2007). Publication bias and merit in ecology. *Oikos*. <https://doi.org/10.1111/j.0030-1299.2007.15686.x>
- Martin, L. J., Blossey, B., & Ellis, E. (2012). Mapping where ecologists work: Biases in the global distribution of terrestrial ecological observations. *Frontiers in Ecology and the Environment*. <https://doi.org/10.1890/110154>
- Mayne, B., Espinoza, T., Roberts, D., Butler, G. L., Brooks, S., Korbie, D., & Jarman, S. (2021). *Non-lethal age estimation of three threatened fish species using DNA methylation: Australian lungfish, Murray cod and Mary River cod: Molecular Ecology Resources*. <https://doi.org/10.1111/1755-0998.13440>
- McKinney, W. (2010). Data structures for statistical computing in Python. In *Proceedings of the 9th Python in science conference* (pp. 56–61). Austin. <https://doi.org/10.25080/majora-92bf1922-00a>
- Meyer, D., Zeileis, A., & Hornik, K. (2023). vcd: Visualizing categorical data. R package. <https://cran.r-project.org/package=vcd>
- Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2010). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *International Journal of Surgery*, 8(5), 336–341. <https://doi.org/10.1016/J.IJSU.2010.02.007>
- Molbert, N., Angelier, F., Alliot, F., Ribout, C., & Goutte, A. (2021). Fish from urban rivers and with high pollutant levels have shorter telomeres. *Biology Letters*, 17(1), 20200819. <https://doi.org/10.1098/rsbl.2020.0819>
- Møller, A. P., & Jennions, M. D. (2001). Testing and adjusting for publication bias. *Trends in Ecology & Evolution*, 16(10), 580–586. [https://doi.org/10.1016/S0169-5347\(01\)02235-2](https://doi.org/10.1016/S0169-5347(01)02235-2)
- Nakagawa, S., Yang, Y., Macartney, E. L., Spake, R., & Lagisz, M. (2023). Quantitative evidence synthesis: A practical guide on meta-analysis, meta-regression, and publication bias tests for environmental sciences. *Environmental Evidence*, 12(1), 1–19. <https://doi.org/10.1186/s13750-023-00301-6>
- O'Dea, R. E., Lagisz, M., Jennions, M. D., Koricheva, J., Noble, D. W. A., Parker, T. H., Gurevitch, J., Page, M. J., Stewart, G., Moher, D., & Nakagawa, S. (2021). Preferred reporting items for systematic reviews and meta-analyses in ecology and evolutionary biology: A PRISMA extension. *Biological Reviews*, 96(5), 1695–1722. <https://doi.org/10.1111/brv.12721>
- Pae, C. U. (2015). Why systematic review rather than narrative review? *Psychiatry Investigation*, 12(3), 417. <https://doi.org/10.4306/PI.2015.12.3.417>
- Perianes-Rodriguez, A., Waltman, L., & van Eck, N. J. (2016). Constructing bibliometric networks: A comparison between full and fractional counting. *Journal of Informetrics*, 10(4), 1178–1195. <https://doi.org/10.1016/J.JOI.2016.10.006>
- Plot, V., Criscuolo, F., Zahn, S., & Georges, J. Y. (2012). Telomeres, age and reproduction in a long-lived reptile. *PLoS ONE*, 7(7), e40855. <https://doi.org/10.1371/JOURNAL.PONE.0040855>
- Polanowski, A. M., Robbins, J., Chandler, D., & Jarman, S. N. (2014). Epigenetic estimation of age in humpback whales. *Molecular Ecology Resources*, 14(5), 976–987. <https://doi.org/10.1111/1755-0998.12247>
- Python Team. (2021). Python programming language. Wilmington, Delaware, United States. <https://www.python.org/>
- R Core Team. (2020). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing.
- Rizopoulos, D. (2006). Irm: An R package for latent variable modeling and item response theory analyses. *Journal of Statistical Software*, 17(5), 1–25.
- Robeck, T. R., Fei, Z., Lu, A. T., Haghani, A., Jourdain, E., Zoller, J. A., Li, C. Z., Steinman, K. J., DiRocco, S., Schmitt, T., & Osborn, S. (2021). Multi-species and multi-tissue methylation clocks for age estimation in toothed whales and dolphins. *Communications Biology*. <https://doi.org/10.1038/s42003-021-02179-x>
- Robeck, T. R., Haghani, A., Fei, Z., Lindemann, D. M., Russell, J., Herrick, K. E. S., Montano, G., Steinman, K. J., Katsumata, E., Zoller, J. A., & Horvath, S. (2023). Multi-tissue DNA methylation aging clocks for sea lions, walruses and seals. *Communications Biology*, 6(1), 359. <https://doi.org/10.1038/s42003-023-04734-0>

- RStudio Team. (2021). RStudio: Integrated development environment for R. Boston, MA. <http://www.rstudio.com/>
- Schwarzer, G., Carpenter, J. R., & Rücker, G. (2015). *Meta-analysis with R*. Springer International Publishing. <https://doi.org/10.1007/978-3-319-21416-0>
- Seabold, S., & Perktold, J. (2010). Statsmodels: Econometric and statistical modeling with python. In *Proceedings of the 9th Python in science conference* (Vol. 57, pp. 10–25080). Austin
- Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, 52(3/4), 591. <https://doi.org/10.2307/2333709>
- Small, H., & Garfield, E. (1985). The geography of science: Disciplinary and national mappings. *Journal of Information Science*, 11(4), 147–159. <https://doi.org/10.1177/016555158501100402>
- Sterne, J. A. C., Egger, M., & Smith, G. D. (2001). Systematic reviews in health care: Investigating and dealing with publication and other biases in meta-analysis. *British Medical Journal*, 323(7304), 101–105. <https://doi.org/10.1136/bmj.323.7304.101>
- Tawfik, G. M., Dila, K. A. S., Mohamed, M. Y. F., Tam, D. N. H., Kien, N. D., Ahmed, A. M., & Huy, N. T. (2019). A step by step guide for conducting a systematic review and meta-analysis with simulation data. *Tropical Medicine and Health*, 47(1), 46. <https://doi.org/10.1186/s41182-019-0165-6>
- Thornton, A., & Lee, P. (2000). Publication bias in meta-analysis: Its causes and consequences. *Journal of Clinical Epidemiology*, 53(2), 207–216. [https://doi.org/10.1016/S0895-4356\(99\)00161-4](https://doi.org/10.1016/S0895-4356(99)00161-4)
- Ujvari, B., Biro, P. A., Charters, J. E., Brown, G., Heasman, K., Beckmann, C., & Madsen, T. (2017). Curvilinear telomere length dynamics in a squamate reptile. *Functional Ecology*, 31(3), 753–759. <https://doi.org/10.1111/1365-2435.12764/SUPPINFO>
- Vernasco, B. J., Dakin, R., Majer, A. D., Haussmann, M. F., Brandt Ryder, T., & Moore, I. T. (2021). Longitudinal dynamics and behavioural correlates of telomeres in male wire-tailed manakins. *Functional Ecology*, 35(2), 450–462. <https://doi.org/10.1111/1365-2435.13715/SUPPINFO>
- Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, 36(3), 1–48.
- Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., & Van Der Walt, S. J. (2020). SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nature Methods*, 17(3), 261–272. <https://doi.org/10.1038/s41592-019-0686-2>
- Webster, J., & Watson, R. T. (2002). Analyzing the past to prepare for the future: Writing a literature review. *MIS Quarterly*, 26(2), xiii–xxiii.
- Wilk, M. B., & Gnanadesikan, R. (1968). Probability plotting methods for the analysis of data. *Biometrika*, 55(1), 1–17. <https://doi.org/10.1093/biomet/55.1.1>
- Wilkinson, G. S., Adams, D. M., Haghani, A., Lu, A. T., Zoller, J., Breeze, C. E., Arnold, B. D., Ball, H. C., Carter, G. G., Cooper, L. N. and Dechmann, D. K. (2021). DNA methylation predicts age and provides insight into exceptional longevity of bats. *Nature Communications*, 12(1), 1615. <https://doi.org/10.1038/s41467-021-21900-2>
- Zhou, P., & Leydesdorff, L. (2010). Fractional counting of citations in research evaluation: An option for cross- and interdisciplinary assessments. Retrieved October 28, 2023, from <http://arxiv.org/abs/1012.0359>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.