



HAL
open science

Are killer bees good for coffee? The contribution of a paper's title and other factors to its future citations

Mark Costello, Karen Beard, Richard Primack, Vincent Devictor, Amanda Bates

► To cite this version:

Mark Costello, Karen Beard, Richard Primack, Vincent Devictor, Amanda Bates. Are killer bees good for coffee? The contribution of a paper's title and other factors to its future citations. *Biological Conservation*, 2019, 229, pp.A1-A5. 10.1016/j.biocon.2018.07.010 . hal-01989800

HAL Id: hal-01989800

<https://hal.science/hal-01989800>

Submitted on 25 Jan 2021

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.

8-8-2018

Are Killer Bees Good for Coffee? The Contribution of a Paper's Title and Other Factors to Its Future Citations

Mark J. Costello
University of Auckland

Karen H. Beard
Utah State University

Richard B. Primack
Boston University

Vincent Devictor
Université Montpellier

Amanda E. Bates
Memorial University of Newfoundland

Follow this and additional works at: https://digitalcommons.usu.edu/eco_pubs



Part of the [Ecology and Evolutionary Biology Commons](#)

Recommended Citation

Mark J. Costello, Karen H. Beard, Richard B. Primack, Vincent Devictor, Amanda E. Bates, Are killer bees good for coffee? The contribution of a paper's title and other factors to its future citations, *Biological Conservation*, 2018, , ISSN 0006-3207, <https://doi.org/10.1016/j.biocon.2018.07.010>.
(<http://www.sciencedirect.com/science/article/pii/S0006320718310000>)

This Article is brought to you for free and open access by the Ecology Center at DigitalCommons@USU. It has been accepted for inclusion in Ecology Center Publications by an authorized administrator of DigitalCommons@USU. For more information, please contact digitalcommons@usu.edu.

Editorial

Are killer bees good for coffee? The contribution of a paper's title and other factors to its future citations

Mark J. Costello, Institute of Marine Science, University of Auckland, Auckland 1142, New Zealand.

Karen H. Beard, Department of Wildland Resources and the Ecology Center, Utah State University, Logan, Utah, 84322-5230 U.S.A.

Richard B. Primack, Biology Department, Boston University, Boston, MA 02215, U.S.A.

Vincent Devictor, Institut des Sciences de l'Evolution, Université Montpellier, CNRS, IRD, 34090, Montpellier, France

Amanda E. Bates, Department of Ocean Sciences, Memorial University of Newfoundland, St. John's A1C 5S7, Canada

Abstract

How can the title of a paper affect its subsequent number of citations? We compared the citation rate of 5,941 papers published in the journal *Biological Conservation* from 1968 to 2012 in relation to: paper length; title length; number of authors; paper age; presence of punctuation (colons, commas or question marks); geographic and taxonomic breadth; the word 'method'; and the type of manuscript (article, review). The total number of citations increased in more recently published papers and thus we corrected citation rate (average number of citations per year since publication) by publication age. As expected, review papers had, on average, twice the number of citations compared to other types of articles. Papers with the greatest geographic or taxonomic breadth were cited up to twice as frequently as narrowly focused papers. Titles phrased as questions, shorter titles, and papers with more authors had slightly higher numbers of citations. However, overall, we found that the included parameters explained only 12% of the variability in citation rate. This suggests that finding a good title is necessary, but that other factors are more important to construct a well-cited paper. We suggest that to become highly cited, a primary requirement is that papers need to advance the science significantly and be useful to readers.

1. Introduction

What features of titles make you want to read the paper? Do you like questions? Do you like titles that announce the key result? Put yourself in the shoes of an editor and other readers. If you were an editor deciding whether a paper was a good fit for your journal, whether it should be sent out for review, or if you want to read beyond the abstract, what features in a title would make you lean a bit more toward reviewing it? And, when you act as a reviewer for papers, what features of a title draw you in, and make you more likely to review it favourably? Would you want to read, or not read, a paper with the title: "Are killer bees good for coffee?"

A paper's title is the first item editors, reviewers, and readers encounter. A well-chosen title creates interest and expectations; a poorly-chosen title suggests that the article may not be worth the time to read. As a result, authors may agonize over how to compose a title that will help their paper get accepted, and then be frequently read and cited afterwards. Here, we review literature on how titles

of papers correlate with their number of citations and test for relationships on number of citations and variables related to the title and paper length, number of authors, the type of paper, and breadth in 5,941 papers published in the journal *Biological Conservation*. Based on our findings, we make recommendations for authors aiming to publish highly cited articles. It should be noted that the titles of these papers may have been improved by going through the review process, and that these studies do not include papers that were rejected.

2. Insights from the literature

2.1. Common findings

Some reviews of the importance of titles for citation rates have found that there are more citations of papers with titles that indicate the results of the paper (Paiva et al. 2012), and papers with broader more conceptual titles (Fox and Burns 2019). Thus, review papers have more citations because of their broader focus (Vanclay 2013, Alimoradi et al. 2016). By contrast, titles indicating a limited scope have fewer citations, and similarly papers with titles that name a geographic region can have fewer citations (Jacques and Sebire 2010, Paiva et al. 2012, Alimoradi et al. 2016). However, methods papers, many very specific to a narrow subject area, dominate the 100 most cited papers of all time (Van Noorden et al. 2014). Previous studies have also found more citations for (a) papers with more authors and for papers with authors from different countries (Leimu and Koricheva 2005, Viera and Gomes 2010, UNESCO 2017), and (b) for longer papers and papers from countries where English is the main language (Leimu and Koricheva 2005). There can also be a significant effect of authors citing other studies within their same country or region (Pasterkamp et al. 2007).

2.2. Contradictory findings

The set of factors influencing citation rates are not consistent across studies. Some studies have found more citations for papers with longer titles (Leimu and Koricheva 2005, Jacques and Sebire 2010, Viera and Gomes 2010, Vanclay 2013), and other studies find more citations with shorter titles (Paiva et al. 2012, Subotic and Mukherjee 2014); while others find no effect of title length (Rostami et al. 2014, Falahati Qadimi Fumani et al. 2015, Alimoradi et al. 2016, Nair and Gibbert 2016). One study found citations increase with the age of the paper (Wang et al. 2013), but another found older papers were cited less (Redner 2004). Two studies found papers with titles that included colons or acronyms to be more cited (Jacques and Sebire 2010, Rostami et al. 2014), but two other studies found these types of punctuation marks, and titles with question marks, to be less cited (Paiva et al. 2012, Nair and Gibbert 2016). Conflicting results reflect the different time periods, numbers of papers analysed, journal impact factor, policies of journals, and perhaps discipline specific patterns (Merrill and Knipps 2014). Certainly, citation rates vary greatly between disciplines (e.g., Viera and Gomes 2010). Here we contribute to this understanding by focusing on citation rates in the field of conservation biology, which has not previously been conducted.

2.3. Expectations for Biological Conservation

Based on past studies, we may thus expect to find higher citation rates for papers that concern methods or are reviews; have more pages or authors; and do not reflect narrow scope, whether geographic or taxonomic. We also may find citations correlated with paper age, title length, or the presence of colons or question marks. Because citation rates are highly influenced by the number of people working in the field, citations rates are best compared within a field of research. Here we compare trends in the nature conservation journal, *Biological Conservation*, established since 1968. Google Scholar has ranked *Biological Conservation* as the top journal in the field of “biodiversity and conservation biology” using the h-index, and it has published over 7,000 papers in 50 years.

3. Methods

We analysed relationships between the mean number of citations each paper accumulated per year (citations rate, as of 6th November 2016) and the age of the publication (i.e., year of publication). We found a strong positive relationship (Figure 1), and we use the residuals of the relationship between double square root transformed citation rate (because the distribution of citation rate is not normally distributed) and year of publication for subsequent analyses. This is described in the legend of Figure 1 and identified as age-corrected citation rate throughout. We test for differences in the age-corrected citation rate with features of the paper's title, number of authors, paper length, and types of papers for 5,941 peer-reviewed articles published in the journal *Biological Conservation* from 1968 to 2012. In total, these articles were cited 239,791 times. More recent years (2013-present) were excluded as these papers would have less citations due to their age, and thus low citation rates. The source data used are freely available at Figshare DOI <https://doi.org/10.17608/k6.auckland.6667622.v1>.

The length of titles and numbers of authors were calculated as the number of characters in the title and author field, respectively. Note for ease of computation this was not the actual number of author surnames. The occurrence of a colon, question mark, and the word 'method' (including methods and methodology), were searched for in all titles using search formulae. In addition, the presence of a geographic or taxonomic name was manually searched for in all titles. Geographic names were scored as: no location; continental or global; super-national, sub-continental, ocean regions; national including Australia, ocean sub-regions and seas; sub-regional: state, province, large lake; or local: national park, city, river. Taxonomic names were scored as: 1 (species or genus), 2 (family), 3 (higher taxonomic group: class, phylum, kingdom), 4 (broad descriptor or group), or 5 (no descriptor or group).

To test for patterns in age-corrected citation rate versus variables indicating the structure and breadth of paper titles, we used a general additive model (GAM) with the package *mgcv* (Wood 2011) in R (Team 2014). This approach allowed fitting of non-linear relationships to detect maximum age-corrected citation rates, if any, for a given variable. The residual structure was visually inspected to ensure that the test assumptions were met. A complicating factor in analysing citation rates is that there are relationships amongst variables. For example, more than half of the top 20 papers with the most authors were review papers. Review papers may also tend to be longer in length than research articles and have more authors, and thus are more highly cited. Relationships amongst the parameters representing variability in the title structure and scope, and citation rate, were produced using the function "pairs" in the raster package (Hijmans, 2016). Reported coefficients and significance levels are for transformed response data that were age-corrected (Figure 1).

4. Results

Paper age, number of authors, paper and title length

Although some old papers were highly cited, there was a strong trend for more recent papers to be more highly cited (Figure 1). Papers were cited at higher rates if they had more authors and pages; see results for "Author length" (positive trend, $p < 0.001$) and "Page count" (positive trend, $p < 0.001$) in Table 1. However, papers with longer titles were cited less; "Title length" (negative trend, $p < 0.001$) Table 1.

Type of paper

Review papers accumulate at least 1.6 times (based on back transforming the coefficients reported in Table 1 for the reference, which were corrected for age) more citations per year in comparison to Articles (e.g., 7.1 versus 4.3 in 2012). Eleven of the top 20 cited papers (adjusted for age) were review papers (Table 2).

Titles with geographic and taxonomic names

Increasing taxonomic breadth was related positively to citation rate (Table 1: wider taxonomic scope, $p < 0.001$ and Figure 2a). For example, in 2012, Articles identifying a particular species or genus by name, accumulated 2.7 citations, while those at the other extreme with no or vague mention of a taxonomic group accumulated 3.7 citations. Increasing geographic breadth also was related positively to citation rate, where titles with broader scope were cited significantly more than titles with less scope (Table 1: Geographic scope (wider), $p < 0.001$ and Figure 2b). In 2012, Articles with titles mentioning a specific region were cited 3.0 times, while those identifying a broad region were cited 4.0 times.

Titles with questions, colons, or stating a result

On average, paper titles that included a question mark were cited 1.3 times as frequently, in comparison to those without. This difference was significant (Table 1, “Question (present)” positive trend, $p = 0.003$, and Figure 2c). Similarly, titles with colons were slightly more cited than those without colons: 1.1 times the number of citations per year (Table 1, “Colon (present)” positive trend, $p = 0.041$, and Figure 2c). There was no significant difference in numbers of citations of papers that included a comma in their title (Table 1, “Comma (present)” $p > 0.05$), or those indicating a ‘method’, versus those that did not (Table 1, “Method (present)” $p > 0.05$).

5. Discussion

As expected, we found Review articles received the most citations. We conclude that papers of more use to other scientists, notably good reviews and papers with broader taxonomic and geographic scope, get more citations. In other words, to have a paper highly cited, scientists should choose the science that advances a general topic rather than being specific to a particular location or species. Thus, we found that papers specific to a species or genus, or a local area, were less cited. The fact that papers with “methods” in their title were not significantly more cited may be because most methods are highly specific to a particular species, habitat or phenomenon, and/or that they do not find wide use. This contrasts with the finding that over all time the top 100 cited science papers were dominated by methods papers (Van Noorden et al. 2014). However, our data set only included 57 methods papers.

Authors may be tempted to try to influence citations by making a title appear to address a general issue when the paper does not. In most journals, this will be noticed by referees and editors who will request the title to be changed appropriately. Alternatively, they may reject the paper because the title does not strongly support the finding indicated in the title. Authors need to frame their work to emphasise its importance and novelty, but not overstate it.

An important caveat in our analysis is that all the papers we analysed have passed review by several referees each, plus scrutiny by at least one editor (Primack 2009). Papers submitted with unclear or misleading titles may either not have made it to peer review (more than half of papers submitted to *Biological Conservation* are immediately rejected by editors), been rejected after peer review, or been revised to improve their title. Once through this review process, we found that whether a title indicated the result of a paper did not correlate with number of citations, with only weak effects of title length, number of authors, page length, and having a comma or question mark. In fact we found that with all the variables included, we could only explain 12% of the variation in citation rate (reported as Deviance in Table 1). These findings are encouraging because they suggest that the underlying science influences the number of citations more than the number of authors, or format and length of a paper’s title.

Our findings contradicted several expectations. Over 44 years, there are some old papers that have never been highly cited while some recent papers have been very highly cited, even though recent

papers have less time to grow in reputation in the literature (Figure 1). The trend for more recent papers to be more highly cited than older ones is probably due to the increasing numbers of researchers working in this field, and the increasing number of journals publishing and articles being written in this research area. Also, editors of leading journals, such as *Biological Conservation*, are becoming more selective, tending to publish articles of wider general interest that are more widely cited.

Despite the lack of a significant relationship between titles that indicate a result, and those that did not, the editors at *Biological Conservation* favour the former. That is because a title indicating a result communicates the key finding of a paper, and indicates the methods and context of the study. For example, here are some possible titles that include a key finding:

OK title: *Relationship of killer bees to fruit set in coffee plants.*

Better (gives the result): *Killer bees increase fruit set in coffee plants.*

Even better (gives the result and has a broad scope so the paper places the results in a general context): *A non-native insect increases fruit set in a tropical crop.*

Also useful would be a review article on this topic: *Pollination systems in tropical agriculture: A review.*

The editors at *Biological Conservation* like titles that accurately reflect what articles are about. Titles set the expectations for papers, so make sure your title accurately reflects the message and content of your paper. As editors, we like to see a title and paper that is aimed squarely at the target audience and content of *Biological Conservation*. The paper should be useful to an international readership interested in research that advances the science and practice of conservation, including the application of conservation principles to natural resource management and policy.

4. Conclusion

So how should you choose a title of your article for *Biological Conservation* that will increase citation rate? First, do research that advances conservation science and management (Primack 2009), and has applications beyond one locality with implications for more than one species. Broadening the research scope is more important than the title of your article. Then make this significant feature clear in the Abstract, Introduction and Discussion. Furthermore, choose a title that is concise, avoids jargon, and indicates the key findings of the study.

Acknowledgements

We thank Karen Foley, Shane Hill, Maggie Hallerud, and Adrienne Holland for assistance in data preparation. Abraham Miller-Rushing provided suggestions on the manuscript.

References

- Alimoradi, F., Javadi, M., Mohammadpoorasl, A., Moulodi, F. and Hajizadeh, M., 2016. The effect of key characteristics of the title and morphological features of published articles on their citation rates. *Annals of Library and Information Studies (ALIS)* 63 (1), 74-77.
- Costello MJ, Beard KH, Corlett RT, Cumming G, Devictor V, Loyola R, Maas B, Miller-Rushing AJ, Pakeman R, Primack RB. 2016. Field work ethics in biological research. *Biological Conservation* 203, 268-271.
- Fox, C.W. and Burns, C.S., 2015. The relationship between manuscript title structure and success: editorial decisions and citation performance for an ecological journal. *Ecology and evolution* 5(10), 1970-1980.
- Falahati Qadimi Fumani, M.R., Goltaji, M. and Parto, P., 2015. The impact of title length and punctuation marks on article citations. *Annals of Library and Information Studies (ALIS)*, 62 (3), 126-132.
- Hijmans RJ (2016). Raster: Geographic Data Analysis and Modeling. R package version 2.5-8. <https://CRAN.R-project.org/package=raster>
- Leimu, R. and Koricheva, J., 2005. What determines the citation frequency of ecological papers? *Trends in Ecology & Evolution* 20 (1), 28-32.
- Merrill, E. and Knipps, A., 2014. What's in a Title? *The Journal of Wildlife Management* 78 (5), 761-762.
- Nair, L.B. and Gibbert, M., 2016. What makes a 'good' title and (how) does it matter for citations? A review and general model of article title attributes in management science. *Scientometrics* 107 (3), 1331-1359.
- Paiva, C.E., Lima, J.P.D.S.N. and Paiva, B.S.R., 2012. Articles with short titles describing the results are cited more often. *Clinics* 67 (5), 509-513.
- Pasterkamp, G., Rotmans, J., de Kleijn, D. and Borst, C., 2007. Citation frequency: A biased measure of research impact significantly influenced by the geographical origin of research articles. *Scientometrics* 70 (1), 153-165.
- Primack, R.B. 2009. Why did we reject your paper? *Biological Conservation* 142, 1559.
- Redner, S., 2004. Citation statistics from more than a century of physical review. *arXiv preprint physics/0407137*.
- Rostami, F., Mohammadpoorasl, A. and Hajizadeh, M., 2014. The effect of characteristics of title on citation rates of articles. *Scientometrics* 98 (3), 2007-2010.
- Subotic, S. and Mukherjee, B., 2014. Short and amusing: The relationship between title characteristics, downloads, and citations in psychology articles. *Journal of Information Science* 40 (1), 115-124.
- Team RC. 2014. R: A language and environment for statistical computing Vienna, Austria. Available from: <http://www.R-project.org> .
- UNESCO 2017. Global Ocean Science Report – the current status of ocean science around the world. UNESCO Publishing, Paris.
- Vanclay, J.K., 2013. Factors affecting citation rates in environmental science. *Journal of Informetrics* 7 (2), 265-271.
- Van Noorden, R., Maher, B. and Nuzzo, R., 2014. The top 100 papers. *Nature News* 514 (7524), p. 550.
- Vieira, E.S. and Gomes, J.A., 2010. Citations to scientific articles: Its distribution and dependence on the article features. *Journal of Informetrics* 4(1), 1-13.
- Wang, D., Song, C. and Barabási, A.L., 2013. Quantifying long-term scientific impact. *Science* 342 (6154), 127-132.

Wood SN. 2011. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society (B)* 73, 3-36.

Table 1. Summary table for generalized additive model results (GAM), including spline fits (s) for publication age-corrected citation rate (as described in Fig. 1b and legend) versus the variables page count, title length and author length), including geographic and taxonomic scope (ordinal variables) and whether the paper included a question, colon, comma, method in the title (no, yes), or was a review. The reference for treatment contrasts was a research article, with geographic and taxonomic scope at the reference of “0”, and question, colon, comma and method equal to “no”. The contrast coefficient estimates are against the reference. edf = estimated degrees of freedom; df = degrees of freedom.

	Coefficient Estimate	Standard Error	t-value	p-value	direction
Reference	-0.660	0.035	-19.010	<0.001	
Page count	0.291	0.018	15.681	<0.001	positive
Geography (wider)	0.012	0.002	7.352	<0.001	positive
Taxonomic scope (wider)	0.027	0.002	11.600	<0.001	positive
Question (present)	0.046	0.015	3.130	0.002	positive
Colon (present)	0.016	0.008	2.043	0.041	positive
Comma (present)	-0.003	0.008	-0.367	0.713	
Method (present)	-0.037	0.035	-1.054	0.292	
Review	0.178	0.018	10.029	<0.001	positive
	edf	df	F-value	p-value	
s (Title length)	2.250	2.613	7.761	<0.001	negative
s (Author length)	2.983	3.000	14.692	<0.001	positive
R-sq.(adj)	0.118				
Deviance	12%				

Table 2. The top 20 papers with the most citations over time based on age-adjusted values (from Figure 1b). * = classified as a review paper in the Scopus database.

Title	Authors	Year	Age-corrected citation rate	Total Citations
The Brazilian Atlantic Forest: How much is left, and how is the remaining forest distributed? Implications for conservation	Ribeiro, M.C. + 4 authors	2009	2.0	971
*Stakeholder participation for environmental management: A literature review	Reed, M.S.	2008	1.8	795
Urbanization as a major cause of biotic homogenization	McKinney, M.L.	2006	1.7	855
*Biodiversity management in the face of climate change: A review of 22 years of recommendations	Heller, N.E. + 1 author	2009	1.6	597
*Global food security, biodiversity conservation and the future of agricultural intensification	Tscharntke, T. + 7 authors	2012	1.5	330
*Genetics and extinction	Frankham, R.	2005	1.5	742
*Does organic farming benefit biodiversity?	Hole, D.G. + 5 authors	2005	1.4	587
*Effects of habitat loss and fragmentation on amphibians: A review and prospectus	Cushman, S.A.	2006	1.3	496
Conservation evaluation and phylogenetic diversity	Faith, D.P.	1992	1.6	1188
Hard choices: Making trade-offs between biodiversity conservation and human well-being	McShane, T.O. + 13 authors	2011	1.2	230
*The fate of Amazonian forest fragments: A 32-year investigation	Laurance, W.F. + 14 authors	2011	1.1	225
*Classification of ecosystem services: Problems and solutions	Wallace, K.J.	2007	1.2	396
*An assessment of the published results of animal relocations	Fischer, J. + 1 author	2000	1.2	567
What to protect? Systematics and the agony of choice	Vane-Wright, R.I. + 2 authors	1991	1.4	874
How much habitat is enough?	Fahrig, L.	2001	1.2	521
*The impacts of roads and other infrastructure on mammal and bird populations: A meta-analysis	Benítez-López, A. + 2 authors	2010	1.0	203
*Ecological functions and ecosystem services provided by Scarabaeinae dung beetles	Nichols, E. + 4 authors	2008	1.0	269
Scattered trees are keystone structures - Implications for conservation	Manning, A.D. + 2 authors	2006	1.0	326
Climate change meets habitat fragmentation: Linking landscape and biogeographical scale levels in research and conservation	Opdam, P. + 1 authors	2004	1.0	373
Faustian bargains? Restoration realities in the context of biodiversity offset policies	Maron, M. + 8 authors	2012	0.9	123

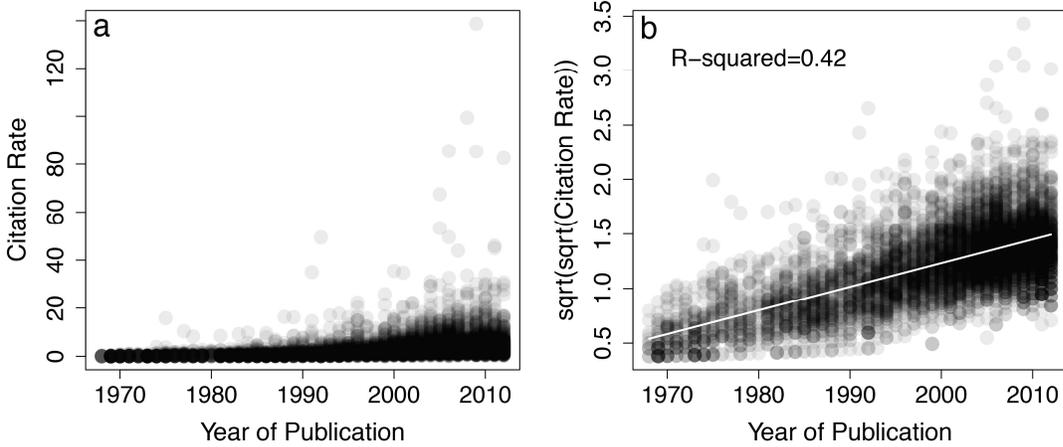


Figure 1. Citations accumulated per year per paper in publications from 1968 to 2012 for (a) raw and (b) transformed response data. The white line is the regression slope and R-squared returned from a linear model. The residuals of the relationship between transformed citation rate and year of publication in (b) are used as the response variable (publication age-corrected) for data visualization across the different predictors and in the statistical model.

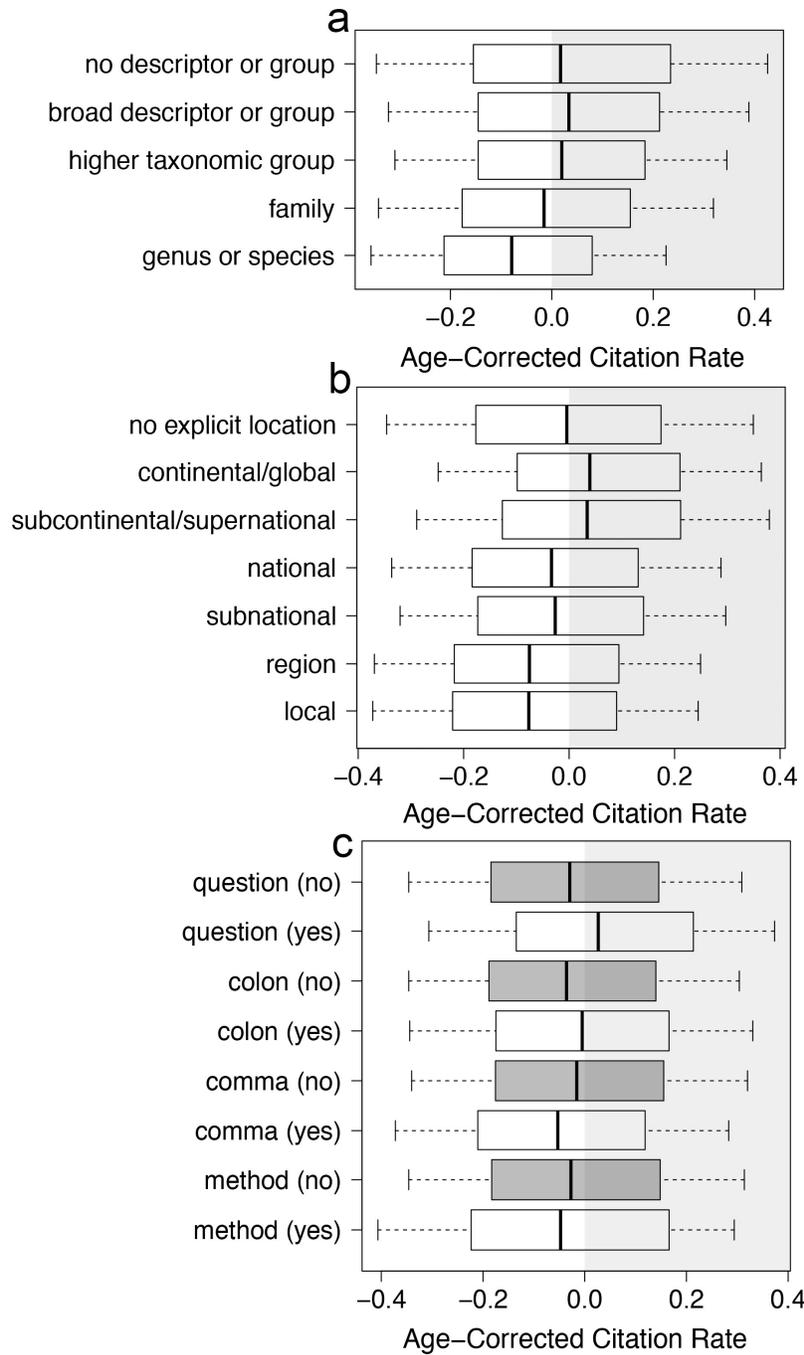


Figure 2. Boxplots of publication age-corrected citation rate (as described in Fig. 1b and legend) for (a) different levels of taxonomic breadth: no descriptor or group ($n = 977$ papers), broad descriptor or group ($n = 819$), higher taxonomic group ($n = 1,395$), family ($n = 328$), and genus or species ($n = 2,422$), (b) different levels of geographic breadth: no explicit location ($n = 2,752$), continental or global ($n = 135$), subcontinental or supranational ($n = 307$), national ($n = 965$), subnational ($n = 557$), region ($n = 413$), local ($n = 812$); and (c) with the word method in their title ($n = 56$) versus without ($n = 5,885$), with ($n = 1,490$) and without ($n = 4,451$) a comma, with ($n = 1,733$) and without ($n = 4,208$) a colon, with ($n = 340$) and without ($n = 5,601$) a question mark in the title. Boxes show range, median (centre line), the top and bottom of the box are the 25th and 75th percentiles. Light gray shading indicates positive age-corrected citation rate values where papers were cited more frequently than the average rate.