STI2022
GRANADA | SEP 7+8+9

26th International Conference on Science and Technology Indicators

"From Global Indicators to Local Applications"

#STI2022GRX

Full paper

STI 2022 Conference Proceedings

Proceedings of the 26th International Conference on Science and Technology Indicators

All papers published in this conference proceedings have been peer reviewed through a peer review process administered by the proceedings Editors. Reviews were conducted by expert referees to the professional and scientific standards expected of a conference proceedings.

Proceeding Editors

Nicolas Robinson-Garcia
Daniel Torres-Salinas
Wenceslao Arroyo-Machado

OPEN ACCESS

Citation: Escabias, M., & Robinson-Garcia, N. (2022). The relation between contribution statements and academic age. In N. Robinson-Garcia, D. Torres-Salinas, & W. Arroyo-Machado (Eds.), *26th International Conference on Science and Technology Indicators*, STI 2022 (sti2279). https://doi.org/10.5281/zenodo.6974732

Collection: https://zenodo.org/communities/sti2022grx/

The relation between contribution statements and academic age[1]

Manuel Escabias[*] and Nicolas Robinson-Garcia[**]

[*] *escabias@ugr.es*
Department of Statistics and Operational Research, University of Granada (Spain)

[**] *elrobin@ugr.es*
EC3 Research Group, University of Granada, Granada (Spain)

Introduction

The allocation of credit has been a long-standing topic in the field of scientometrics and research evaluation. The increase of multi-authored papers was already an issue of concern in the 1980s when Price (1981) proposed fractionalizing co-authorships in order to avoid author inflation or hyperauthorship (Cronin, 2001).

The continuing rise of collaboration and team science challenges traditional notions of authorship (Fontanarosa et al., 2017) and raise the question on how to adapt counting methods to different collaboration and co-authorship practices. Since De Solla Price's first proposal on fractional counting, numerous methods have been proposed which vary on the approach (i.e., full, fractionalized counting) and the methodology proposed (Gauffriau, 2021 offers a review on counting methods while Leydesdorff & Park, 2016 discuss different approaches for fractional counting when producing co-authorship matrices).

Others have called for an abandonment of the idea of authorship as credit and move towards the concept of contributors (Allen et al., 2014). Researchers are assumed to be partly responsible of the research output, distributing tasks and specializing in specific activities. This perspective goes in hand with the idea of scientific work as a distributed and collaborative activity (Larivière et al., 2016) in which researchers specialize on conducting certain tasks (Robinson-Garcia et al., 2020b). In this way, many journals integrate both perspectives (especially in the biomedical fields) by including, along with the list of authors, contribution statements in which they specify the role played by each of them.

Despite the introduction of initiatives like the Contributor Roles Taxonomy (CRediT), adopted now by many journals (Allen et al., 2019; Larivière et al., 2020), research evaluation practices still rely on the notion of authorship and beyond the existence of a few scientometric studies analyzing these contribution statements, no real application can be observed in terms of implementation in hiring, promotion, selection or other assessment exercises in academia.

[1] This work is part of the R&D project COMPARE (REF: PID2020-117007RA-I00) supported by the Spanish Ministry of Science and Innovation (MCIN/ AEI/10.13039/501100011033/). More information on the project available at http://compare-project.eu. Nicolas Robinson-Garcia is currently supported by a Ramón y Cajal Fellowship granted by the Spanish Ministry of Science and Innovation (REF: RYC2019-027886-I).

Evaluation policies tend to assume a relation between author order and the contributions made by researchers. First, last and corresponding authors and normally assumed to contribute the most (Chinchilla-Rodríguez et al., 2019; Moya-Anegón et al., 2013). But there is no empirical basis for this assumption (Robinson-Garcia et al., 2020b; Sauermann & Haeussler, 2017). Furthermore, approaches promoting leadership based on author order have perversive effects on research careers and the production of knowledge. On the one hand, they corrupt and expand ill-practices in academia (Jabbehdari & Walsh, 2017). On the other hand, they relegate and negatively affect the career of those who cannot opt to gain leading positions in author order (Milojević et al., 2018), affecting the composition of the scientific workforce.

Tscharntke et al. (2007) discuss the different criteria used on author order and how these criteria can affect individual level research assessments. Among others, they point at seniority as being a factor that may confound the relation between author order and division of labour. In a different study, Corrêa Jr. et al. (2017) demonstrate an asymmetrical relationship between the total number of authors and the number of contributions exerted by each of them. In this study we go a step further and analyse the relation between academic status and contributorship, and variations depending on the total number of authors identified. The aim is to understand how academic seniority relates to the task distribution of authors.

Data and methods

This paper uses an existing dataset (Dataset 1) of publications and contribution statements derived from PLOS journals (Robinson-Garcia et al., 2020a). This dataset contains 85,260 distinct PLOS papers published during the 2006-2013 period. It combines bibliometric and contributorship information and it consists of author-publication combinations. For each author-publication combination up to 7 contribution statements are included. More information on the collection of the data is available at Larivière et al. (2016), while (Robinson-Garcia et al., 2020b) provides further details on how bibliometric variables from Web of Science were extracted and computed. Table 1 includes a definition of the variables used in this study and the source from which they were originally extracted.

Following we describe further modifications made to the dataset. Author-publication rows in which authors reported an academic age above 50 years were removed as well as publications in which at least one author did not report contributing at all (probably due to data extraction errors). The final dataset comprises a total of 72,315 distinct publications.

The goal of this study is to understand how researchers' age influences the type of tasks the conduct, depending on the position they hold in the author order. We focus on these two variables based on the findings by (Sauermann & Haeussler, 2017), who noted 'that authorship and contribution disclosures not only reflect objective contributions but also are shaped by important social dynamics' (p. 9). Academic age can serve as a proxy for academic status which can greatly influence the types of tasks expected based on academics' career stage.

For this, we fit a multinomial-response model. This regression model is used when the dependent variable is categorical. The dependent variable is factor or a matrix with K columns, which will be interpreted as counts for each of K classes. The model provides the probability of conducting a given contribution as opposed to one which is used as a reference, that is,

$$P\{C_i|C_r\}$$

Where Ci is one of the five possible contributions reported and Cr is the contribution used as reference. As authors may conduct more than one task, we need to adapt the dataset by assuming that each author-contribution combination is conducted independently.

Table 1. Definition of variables

Variable	Description	Source
Pub ID	Digital Object Identifier assigned to the publication	PLOS
Researcher ID	Created using a name disambiguation algorithm which clusters publications based on rule-based scorings	Caron & van Eck (2014)
Author order	Position of the author in a publication	Web of Science
Number of authors	Total number of authors of a publication	Web of Science
Number of countries	Total number of countries to which authors are affiliated	Web of Science
Current academic age	Years passed since first publication based on the data of a given publication	(Nane et al., 2017)
Contribution	Five types of contributions are considered: wrote the paper (WR), analyzed the data (AD), performed the experiments (PE), Conceived and designed the experiments (CE) and contributed reagents/materials/analysis tools (CT)	PLOS

The analysis was conducted using the R software (R Core Team, 2021), and the R package nnet (Ripley & Venables, 2022). Analyses are conducted based on the number of authors per paper in order to account for differences on division of labour dependent on team size. In this paper we include findings for papers co-authored by 2, 3 and 4 researchers.

Results and discussion

Regression analyses

Table 2 shows the distribution of researchers based on the share of contributions they claim to have conducted. Already in this table some differences are observed based on the age of academics. We observe that the proportion of researchers who report writing the manuscript and performing the experiments drops as researchers become more senior. In the latter case, this drop is quite sharp going from over 30% of academics in their junior stage (< 5 years since their first publication) to around 7% for senior researchers (those who published their first paper at least 30 years ago). For the rest of the contributions the share of researchers who report conducting them increase based on their age, with the biggest increase (8.5%) found for analysing the data.

Table 2. Share of researchers conducting each contribution by academic age

Academic age	WR	AD	PE	CE	CT
< 5 years	23.4	15.9	30.4	13.1	17.2
5-15 years	22.4	22.0	19.1	15.2	21.4
15-30 years	20.8	26.0	10.3	18.7	24.2
≥ 30 years	20.1	27.4	7.2	19.6	25.7

Tables 3 shows the estimated parameters of the multinomial-response regression models when subsetting for papers with 2 authors. All variables are statistically significant (p-values

> 0.005), meaning that the observed relation in Table 2 between academic age and type of contribution does exist.

Table 3. Multinomial-response regression models based on papers authored by 2 researchers

	(Intercept)		Age		Author order [b]	
Contribution [a]	Coef	Std dev	Coef	Std dev	Coef	Std dev
P{CE|AD}	-0.14*	0.02	0.01*	0.00	0.27*	0.04
P{PE|AD}	0.04	0.02	-0.02*	0.00	-0.53*	0.04
P{CT|AD}	-0.67*	0.03	0.01*	0.00	0.32*	0.04
P{WR|AD}	-0.05*	0.02	0.01*	0.00	0.26*	0.04

[a] The reference category is *Analysed the data* [b] The reference category is 1st author * P < 0.05Residual Deviance: 126215.7
AIC: 126239.7

In Table 4 we show the results when papers are co-authored by three researchers. In this case, the model cannot explain the contributions CE, PE and WR based on age and position when authors are in a middle position. Finally, in Table 5 we show the results for papers with 4 co-authors, in this case grouping author order by first, middle and last author. In this case, again all variables seem to be statistically significant.

Table 4. Multinomial-response regression models based on papers authored by 3 researchers

	(Intercept)		Age		author order [b] (2nd)		Author order [b] (3rd)	
Contribution [a]	Coef	Std dev	Coef	Std dev	Coef	Std dev	Coef	Std dev
P{CE|AD}	-0.19*	0.02	0.02*	0.00	-0.02	0.03	0.20*	0.03
P{PE|AD}	0.05*	0.02	-0.03*	0.00	0.01	0.03	-0.51*	0.03
P{CT|AD}	-0.85*	0.02	0.01*	0.00	0.37*	0.03	0.29*	0.03
P{WR|AD}	-0.10*	0.02	0.02*	0.00	-0.03	0.03	0.17*	0.03

[a] The reference category is *Analysed the data* [b] The reference category is 1st author * P < 0.05Residual Deviance: 260548.3
AIC: 260580.3

Table 5. Multinomial-response regression models based on papers authored by 4 researchers differentiating between first, middle and last author

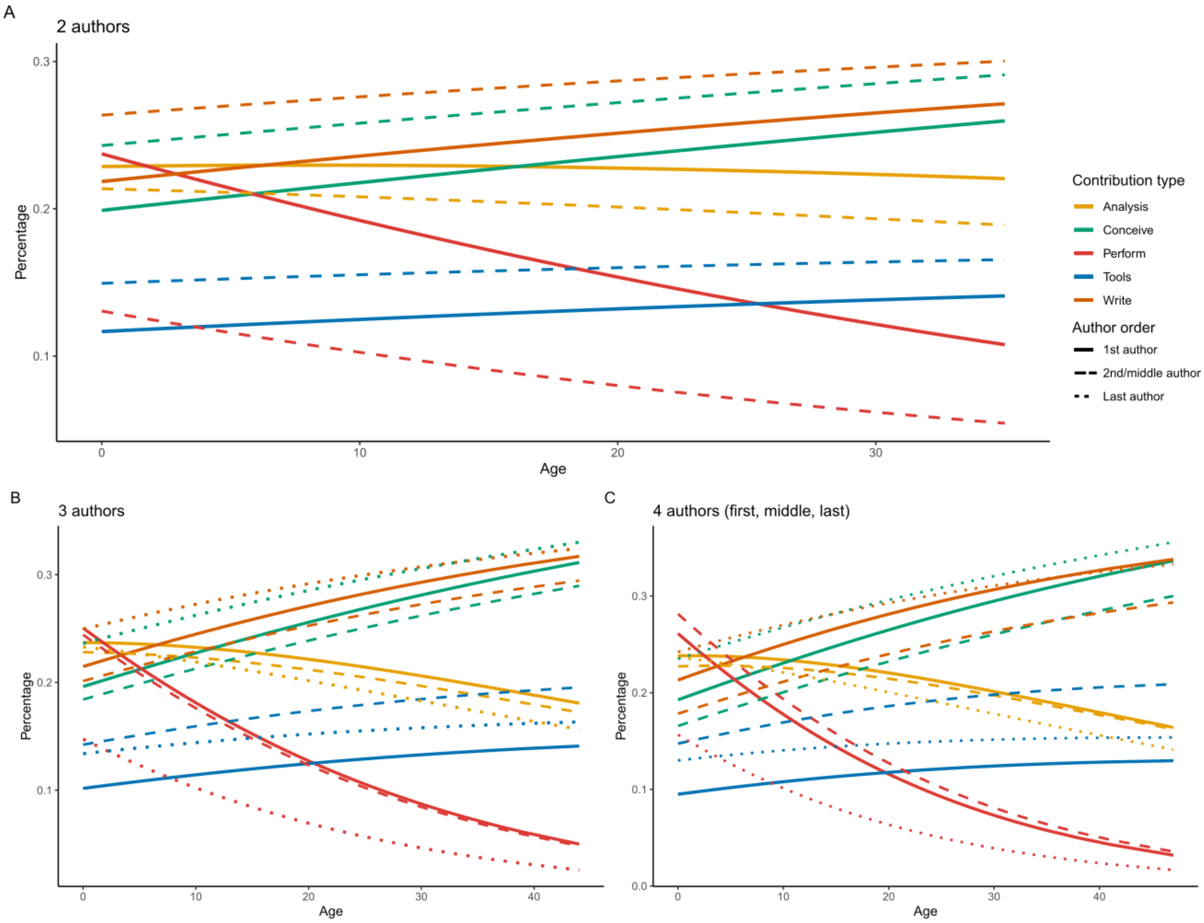
	(Intercept)		Age		Author order (2nd & 3rd)		Author order (4th)	
Contribution [a]	Coef	Std dev	Coef	Std dev	Coef	Std dev	Coef	Std dev
P{CE|AD}	-0.21*	0.02	0.02*	0.00	-0.10*	0.02	0.21*	0.03
P{PE|AD}	0.09*	0.02	-0.04*	0.00	0.12*	0.02	-0.51*	0.03
P{CT|AD}	-0.92*	0.02	0.01*	0.00	0.49*	0.03	0.32*	0.03
P{WR|AD}	-0.11*	0.02	0.02*	0.00	-0.13*	0.02	0.14*	0.03

[a] The reference category is *Analysed the data* * P < 0.05
Residual Deviance: 353361.4
AIC: 353393.4

Relation between age and contributorship by number of authors per paper

Figure 1 shows the predicted probabilities of researchers for conducting each of the five contributions based on their age and position for papers with 2 authors (Fig 1A), 3 authors (Fig 1B) and 4 authors (Fig 1C). The first thing to note is the low probability assigned for all contribution statements, never going beyond 0.4. This results on a model which with very low predictive capacity.

Figure 1. Probability curves of reporting a given contribution based on academic age for publications with (A) 2 authors, (B) 3 authors and (C) 4 authors



Still, the differences observed show that the model does allow to explain influences. For instance, we clearly observe that tasks such as writing and conceiving and designing the experiments tend to be related to last authors first, and secondly to first authors. Furthermore, performing the experiments seems to be a task more related with age rather than author position, showing a step decline as researchers are more senior. This is also observed, but as dramatic, for analysing the data. Regarding contributing with tools, this variable is relatively stable regardless of the age, showing a higher probability of conducting it for second or middle author. However, we do observe a slight increase over time which may be related with the role of assisting and supporting junior researchers as noted elsewhere (Robinson-Garcia et al., 2020b).

Concluding remarks

This papers builds on previous work (Robinson-Garcia et al., 2020b, 2021) in which we try to characterize the diversity of roles researchers play in their work life. Specifically, this paper relates to the type of contributions they conduct when collaborating with other researchers on the production of scientific outputs. For this, we use a dataset which combines contribution statements along with bibliometric data on researchers' academic age and position in the author order. With this, we explain how age and author order affect the probability of researchers on conducting each of the contribution statements using multinomial-responsive regression modelling.

In this paper we focus solely on papers authored by up to 4 researchers. These models are not good for predicting contributorship. Something which as already noted elsewhere (Sauermann & Haeussler, 2017). Still, they do show that differences on contributions are not only explained by author order but also by seniority. Next, we plan to continue expanding these analyses to papers conducted by larger teams, as well as to researchers working in areas other than biomedical fields.

References

Allen, L., O'Connell, A., & Kiermer, V. (2019). How can we ensure visibility and diversity in research contributions? How the Contributor Role Taxonomy (CRediT) is helping the shift from authorship to contributorship. *Learned Publishing*, *32*(1), 71–74. https://doi.org/10.1002/leap.1210

Allen, L., Scott, J., Brand, A., Hlava, M., & Altman, M. (2014). Publishing: Credit where credit is due. *Nature*, *508*(7496), 312–313. https://doi.org/10.1038/508312a

Caron, E., & van Eck, N. J. (2014). Large scale author name disambiguation using rule-based scoring and clustering. *19th International Conference on Science and Technology Indicators.“Context Counts: Pathways to Master Big Data and Little Data”* (pp. 79–86). http://www.researchgate.net/profile/Tindaro_Cicero/publication/265396216_Research_quality_characteristics_of_publications_and_socio-demographic_features_of_Universities_and_Researchers_evidence_from_the_Italian_VQR_2004-2010_evaluation_exercise/links/540d89180cf2df04e754b658.pdf#page=91

Chinchilla-Rodríguez, Z., Sugimoto, C. R., & Larivière, V. (2019). Follow the leader: On the relationship between leadership and scholarly impact in international collaborations. *PLOS ONE*, *14*(6), e0218309. https://doi.org/10.1371/journal.pone.0218309

Corrêa Jr., E. A., Silva, F. N., da F. Costa, L., & Amancio, D. R. (2017). Patterns of authors contribution in scientific manuscripts. *Journal of Informetrics*, *11*(2), 498–510. https://doi.org/10.1016/j.joi.2017.03.003

Cronin, B. (2001). Hyperauthorship: A postmodern perversion or evidence of a structural shift in scholarly communication practices? *Journal of the American Society for Information Science and Technology*, *52*(7), 558–569. https://doi.org/10.1002/asi.1097

Fontanarosa, P., Bauchner, H., & Flanagin, A. (2017). Authorship and Team Science. *JAMA*, *318*(24), 2433–2437. https://doi.org/10.1001/jama.2017.19341

Gauffriau, M. (2021). Counting methods introduced into the bibliometric research literature 1970 – 2018: A review. *Quantitative Science Studies*, 1–52. https://doi.org/10.1162/qss_a_00141

Jabbehdari, S., & Walsh, J. P. (2017). Authorship Norms and Project Structures in Science. *Science, Technology, & Human Values*, 0162243917697192. https://doi.org/10.1177/0162243917697192

Larivière, V., Desrochers, N., Macaluso, B., Mongeon, P., Paul-Hus, A., & Sugimoto, C. R. (2016). Contributorship and division of labor in knowledge production. *Social Studies of Science*, *46*(3), 417–435. https://doi.org/10.1177/0306312716650046

Larivière, V., Pontille, D., & Sugimoto, C. R. (2020). Investigating the division of scientific labor using the Contributor Roles Taxonomy (CRediT). *Quantitative Science Studies*, 1–18. https://doi.org/10.1162/qss_a_00097

Leydesdorff, L., & Park, H. W. (2016). Full and Fractional Counting in Bibliometric Networks. *ArXiv:1611.06943. Retrieved from* https://doi.org/10.48550/arXiv.1611.06943

Milojević, S., Radicchi, F., & Walsh, J. P. (2018). Changing demographics of scientific careers: The rise of the temporary workforce. *Proceedings of the National Academy of Sciences*, *115*(50), 12616–12623. https://doi.org/10.1073/pnas.1800478115

Moya-Anegón, F., Guerrero-Bote, V. P., Bornmann, L., & Moed, H. F. (2013). The research guarantors of scientific papers and the output counting: A promising new approach. *Scientometrics*, *97*(2), 421–434. https://doi.org/10.1007/s11192-013-1046-0

Nane, G. F., Larivière, V., & Costas, R. (2017). Predicting the age of researchers using bibliometric data. *Journal of Informetrics*, *11*(3), 713–729. https://doi.org/10.1016/j.joi.2017.05.002

Price, D. de S. (1981). Multiple Authorship. *Science*, *212*(4498), 986-986. https://doi.org/10.1126/science.212.4498.986.b

R Core Team. (2021). R: A Language and Environment for Statistical Computing (4.1.2) [Computer software]. R Foundation for Statistical Computing. http://www.r-project.org/

Ripley, B., & Venables, W. (2022). nnet: Feed-Forward Neural Networks and Multinomial Log-Linear Models (7.3-17) [Computer software]. https://CRAN.R-project.org/package=nnet

Robinson-Garcia, N., Costas, R., Nane, T., & Leeuwen, T. N. van. (2021). Valuation regimes in academia: Researchers' attitudes towards their diversity of activities and academic performance. *SocArXiv*. Retrieved from https://doi.org/10.31235/osf.io/ve7d3

Robinson-Garcia, N., Costas, R., Sugimoto, C. R., Larivière, V., & Nane, G. F. (2020a). *Datasets on contributorship and bibliometric variables for the study 'Task specialization across research careers'* [Data set]. Zenodo. Retrieved from https://doi.org/10.5281/zenodo.3891055

Robinson-Garcia, N., Costas, R., Sugimoto, C. R., Larivière, V., & Nane, G. F. (2020b). Task specialization across research careers. *ELife*, *9*, e60586. https://doi.org/10.7554/eLife.60586

Sauermann, H., & Haeussler, C. (2017). Authorship and contribution disclosures. *Science Advances*, *3*(11), e1700404. https://doi.org/10.1126/sciadv.1700404

Tscharntke, T., Hochberg, M. E., Rand, T. A., Resh, V. H., & Krauss, J. (2007). Author Sequence and Credit for Contributions in Multiauthored Publications. *PLOS Biology*, *5*(1), e18. https://doi.org/10.1371/journal.pbio.0050018