



Author contributions and allocation of authorship credit: testing the validity of different counting methods in the field of chemical biology

Pär Sundling¹

Received: 8 April 2022 / Accepted: 2 March 2023
© The Author(s) 2023

Abstract

This paper explores the relationship between an author's position in the bylines of an article and the research contributions they have made to analyze the validity of five bibliometric counting methods (arithmetic, fractional, geometric, harmonic, and harmonic parabolic author credit allocation) in the field of Chemical Biology. By classifying the tasks found in the author contribution statements of articles published in Nature Chemical Biology according to a three-tiered scheme, it was possible to divide the authors into three types: core-layer authors, middle-layer authors, and outer-layer authors. When ordering the authorships according to the position in the bylines, there is a distinct u-shaped distribution for the share of authors involved in writing the paper or designing the research (i.e., core authors) and for the average number of tasks performed by each author. The harmonic parabolic model best distributes author credit according to the observed empirical data. It also outperforms the other models in predicting which authors are core authors and which are not. The harmonic parabolic model should be the preferred choice for bibliometric exercises in chemical biology and fields with similar practices regarding authorship order.

Keywords Authorship credit · Author contributions · Validity · Counting methods · Byline hierarchy

Introduction

Over the last two decades, journals in different fields have increasingly required authors to disclose their contributions as part of the research paper (Larivière et al., 2021). As authorship is a proxy for scientific productivity (Cronin, 2001), coupled with an increasing trend to institute metric-guided evaluative mechanisms for governing academia (e.g., Abbott et al., 2010; Hicks, 2012; Wilsdon et al., 2015), there is a need to validate different models for the allocation of author credit against the information presented in such author contribution statements. Although Hagen (2010, 2013) has evaluated several models for

✉ Pär Sundling
par.sundling@umu.se

¹ Department of Sociology, Inforsk, Umeå University, Umeå, Sweden

allocating author credit against perceived author credit scores, no study to date has validated such models against the author contribution statements of scientific articles. This paper aims to address this issue.

It is evident from the literature (Egghe et al., 2000; Gauffriau & Larsen, 2005; Gauffriau et al., 2008; Huang et al., 2011; Vavryčuk, 2018) that the use of different counting procedures has bearings on the number of publications and citations assigned to authors. For example, using fractional counting and thus disregarding taking into account the information present in the structure of the byline introduces an equalization bias (Hagen, 2013, 2014a, 2015). It has a strong distortional effect that is "inevitably compounded in bibliometric indices and performance rankings" (Hagen, 2014b, p. 626). The routine practice of determining author credit by dividing one credit equally among all coauthors of a research paper is inappropriate if the alphabet does not purposely structure the bylines (Hagen, 2013). A majority of the alternative models for determining author credit are, in their basic form, rank-dependent, meaning they share the basic assumption that position in the author bylines is apportioned according to the weight of contributions and structured in descending order. In cases where the least important authors are positioned somewhere in the middle, such models become problematic and risk overestimating some authors' credit and underestimating others' credit.

Allocation models vary in their assumptions about the logic underpinning author order. For example, fractional counting assumes that each author has contributed equally and apportions credit conversely, while harmonic counting assumes a descending ordering logic with regard to the authors' relative contributions. Several bibliometric papers drawing upon the author contributions statements provided by journals report that first and last authors typically contribute to more tasks than middle authors do (Baerlocher et al., 2007; Larivière et al., 2016, 2021; Sauermann & Haeussler, 2017; Sundling, 2017; Yang et al., 2017). Several studies based on interviews and surveys also noted that middle authors contribute less than first and last authors (e.g., Louis et al., 2008; Shapiro et al., 1994; Wren et al., 2007). These results indicate that an allocation model should give more credit to the first and last author and less to the authors between these positions.

However, there are several reasons to question the grouping together of authors between first and last. First, second authors contribute more than authors closer to the middle and end of the byline order (Baerlocher et al., 2007; Wren et al., 2007). Second, the increase in team size and coauthored articles in science (Larivière et al., 2015; Wuchty et al., 2007) comes with an increase in the number of primary and supervisory authors on the same paper (Mongeon et al., 2017), some of which are bound to end up between the first and last positions. Third, and likely partly as a consequence of "team science," there is an increase in the number of scientific articles in which two or more authors claim "equal co-first authorship" (Akhbue & Lautenbach, 2010; Hosseini & Bruton, 2020) or appear as corresponding authors (Hu, 2009). All in all, treating middle authors as one heterogeneous group conceals significant differences in this seemingly diverse group of authors. This also means that the question of how a model for allocating authorship credit should best apportion credit to authors between first and last is still an open question.

Scientific fields differ in aspects such as the degree of collaborative activity, authorship practices, how research is organized, and publication practices. In high energy physics, the bylines are regulated by a labor approach in which everyone working on a project ends up on an alphabetically ordered author list (Birnholtz, 2006; Knorr-Cetina, 1999). Mathematics and Economics also exhibit a high degree of byline alphabetization (Frandsen & Nicolaisen, 2010; Waltman, 2012), while in biomedicine and clinical research, the bylines tend to be structured according to a principle by which the primary authors are first, and the

supervisory authors are last (Mongeon et al., 2017). As author inclusion and byline order differ per field, so will the recommended author credit allocation model. This paper is to be considered a case study in the chemical biology field.

This paper explores the relationship between an author's position in the bylines of a research article and the research contributions they have made in order to analyze the validity of five bibliometric counting methods in Chemical Biology. The research questions explored in this paper are:

- Research question 1: What logic structures the byline ordering in Chemical Biology with regard to the authors contributions?
- Research question 2: How well do models for allocating author credit reflect the number of contributions made by authors in Chemical Biology?
- Research question 3: How well do models for allocating author credit predict core authors in Chemical Biology?

The field of chemical biology can be considered a part of the lab-oriented life sciences. Scientists working within this field often have a background in chemistry or biology (or, on some occasions, are specially trained chemical biologists) and bring together a diverse array of experimental techniques and theoretical knowledge (Ostler, 2007).

Theory

Ordering the bylines

Tscharntke et al. (2007) present four types of name ordering:

1. The sequence-determines-credit approach (SDC), wherein the sequence of authors reflects the usually declining importance of their contribution.
2. The equal contribution approach (EC), wherein authors use alphabetical sequence to reflect having contributed equally to the research. According to Waltman (2012, p. 704), this is more common in fields where the average author team is either small or large, while in fields such as the Medical and Life Sciences, "intentional alphabetical authorship is a virtually non-existent phenomenon." However, in biomedicine and clinical medicine, partially structuring the middle of the bylines according to the alphabet is increasing (Mongeon et al., 2017). Sometimes structuring the bylines according to the alphabet is only a custom and is not meant to convey any information regarding the contributions of authors (Egghe et al., 2000).
3. The first-last-author-emphasis approach (FLAE), wherein the first and last authors have made the most significant contributions. This practice is well-established in many labs and scientific fields. Statements regarding equal contributions or the presence of one or more corresponding authors are also versions of this approach.
4. The percent-contribution-indicated approach (PCI), wherein a percentage score details each author's contribution. I argue that expressing the authors' contributions in an article's author contribution statement can be regarded as a qualitative version of the PCI approach. Even though there is no percentage score, it is usually quite clear who has contributed in a major or a minor way.

The approaches mentioned above are sometimes combined, for example, by partially structuring the bylines by alphabet to indicate equal contributions by some, but not all, of the authors (Mongeon et al., 2017; Waltman, 2012).

Models for allocating authorship credit

Four models for allocating authorship credit are especially prominent in the bibliometric literature (see Waltman (2016, Section 7) and Xu et al. (2016) for an overview): (1) fractional allocation (Price, 1981); (2) harmonic allocation (Hodge & Greenberg, 1981); (3) geometric allocation (Egghe et al., 2000); and (4) arithmetic allocation (Van Hooydonk, 1997). Hagen (2010, p. 792) tests these four empirically, concluding that harmonic allocation “provides unrivalled accuracy, fairness and flexibility”. One should note that these four allocation models are not the only ones (e.g., Assimakis & Adam, 2010; Lukovits & Vinkler, 1995; Stallings et al., 2013; Trueba & Guerrero, 2004). Many papers also deal specifically with author credit allocation in the construction of the Hirsch-index and its variants (e.g., Jian & Xiaoli, 2013) or include it as an essential part of constructing weighting schemes for ranking scientific articles (e.g., Zhang et al., 2019) or researchers (e.g., Vavryčuk, 2018). Several allocation models proposed by the referenced papers above are tested empirically in Hagen (2013), again with the harmonic allocation model being the most accurate.

This paper analyzes five allocation models. The first is harmonic allocation, which is included based on the performances relative to other models mentioned above. The second is fractional allocation, which is included because it is the most implemented approach, besides giving full credit to all authors of a publication. Geometric and Arithmetic allocation is also included to compare the results with earlier studies of allocation models. The formulas for these first four allocation models are:

1. Fractional *i*th author credit = $\frac{1}{n}$
2. Harmonic *i*th author credit = $\frac{\frac{1}{i}}{1 + \frac{1}{2} + \dots + \frac{1}{n}}$
3. Geometric *i*th author credit = $\frac{2^{n-i}}{2^n - 1}$
4. Arithmetic *i*th author credit = $\frac{n+1-i}{(1+2+\dots+n)}$

By denoting *i* as the index for author position and *n* as the total number of authors on the publication, it is possible to calculate the credit score for each author of a paper.

All formulas above provide additive weights, meaning that they sum to 1 and therefore do not inflate the total publication count. While fractional allocation is rank-independent, the remaining three are rank-dependent; for fractional allocation, all authors get an equal share, while the other models give the first author more than the second and the second more than the third (and so on). In this way, fractional allocation is the model best suited for bylines structured according to the EQ approach, while the other three models are best suited for the SDC approach. If the SDC and FLAE approach structure the bylines, then it is possible to modify the calculations to give equal credit to the first and last author, as described in Hagen (2010) and Liu & Fang (2012). The same modification procedure can be extended to include all corresponding authors, and all authors stated to have contributed equally.

The fifth allocation model to be included is one proposed by Aziz & Rosing (2013), but they did not provide a name for this weighing algorithm. For the sake of simplicity,

this paper calls it the harmonic parabolic model for allocating authorship credit. There are two reasons for including it. First, it represents a radically different way of interpreting the byline hierarchy when compared to traditional allocation models. The basis for this model is that most credit should be given to the first and last author and decrease from those positions so that the median position gets the least credit. Moreover, it has not been tested empirically before. The formula for calculating it is:

$$\text{Harmonic parabolic } i\text{th author credit} = \frac{1+|n+1-2i|}{\frac{1}{2}n^2+n(1-D)}$$

By denoting i as the index for author position and n as the total number of authors on the publication, with $D = 0$ if n is an even number or $D = \frac{1}{2n}$ if n is an odd number, it is possible to calculate the score for each author of a paper. The harmonic parabolic model is additive, rank-dependent, and theoretically best suited to use when a combination of the SDC and FLAE approach structures the bylines. That is, in scientific fields where the distribution of key contributors is u-shaped.

Method

The point of departure for this study is the research papers published in Nature Chemical Biology in 2013 and 2014. Each such paper (being either an article or a brief communication) comes with an author contribution statement that “specifies the contribution of every author” (“Authorship: authors & referees @ npg,” 2015). The journal requires the authors to include such a statement, but they are free to structure it and be as detailed as they see fit. See the Appendix for an example of an author contribution statement.

Creating a data set

Bibliographic data concerning all published documents in Nature Chemical Biology for 2013 and 2014 were downloaded from Web of Science (WOS). By using the Digital Object Identifiers (DOI) in the WOS data, it was possible to harvest full-text XHTML data for each document from the journal home page. Using a script written in Tool Command Language (Tcl), the full-text XHTML data of each document was parsed for author contribution statements and, if found, were added to the bibliographic data for the specific document. All documents that did not include author contribution statements were discarded from the data set—this meant only keeping documents of the article and brief communication type. Three additional documents had to be discarded due to errors in the author contribution statement. All documents ($n = 14$) exhibiting hyperauthorship, which is defined by Morris & Goldstein (2007, p. 1766) as “any article with 20 or more authors”, were excluded from the dataset. There were too few papers with an author team of 18 ($n = 2$) and 19 ($n = 1$) authors to warrant inclusion in the final data set, leaving 208 research papers for final analysis. There were at least six papers representing all the other author team sizes. Each paper resulted from collaborative activity (i.e., they were coauthored).

Each author contribution statement was manually scanned for descriptions of authors performing specific work tasks. For each task performed by an author, an item was created in an authorship-task database (if several authors were registered in the author contribution statement of a paper as performing the task, an item was created for each authorship-task combination). In order to classify the tasks in the authorship-task database, we followed

the classification procedure used in Baerlocher et al. (2007): classification, pilot-testing; subject expert review; modification; and reclassification. A preliminary analysis was thus made after classifying all tasks according to a three-layered taxonomy initially developed by Davenport & Cronin (2001) and further operationalized for the classification of author contribution statements by Danell (2014). After discussions with a chemical biologist, the taxonomy and the way specific tasks were classified were modified. The data were then reclassified using the modified taxonomy shown in Table 1. A total of 4955 entries in the authorship-task database were classified according to this taxonomy. The modification and testing of the taxonomy—the intercoder reliability rate¹ was 91,6%—is further detailed in Sundling (2017).

From the entries in the authorship-task database, an authorship database was created wherein each authorship was represented by only one item and classified according to a three-step procedure: First, I classified all authors performing at least one core-layer task as core-layer authors; Second, authors performing at least one middle-layer task but no core-layer tasks were classified as middle-layer authors; Third, all authors not performing any core or middle-layer tasks were classified as outer-layer authors. A total of 1743 entries (authorships) in the authorship database were classified according to this procedure.

While the terms author and authorship have different meanings in bibliometrics, they are both used in this text (for readability) to indicate authorship. Specifically, if one author performs tasks on several papers, they are represented by several items in the authorship database.

Calculating credit scores: the standard version and the special version

Credit scores were calculated for each author in the authorship database according to the following models for allocating authorship credit: fractional allocation, harmonic allocation, geometric allocation, arithmetic allocation, and harmonic parabolic allocation (see Section “[Models for allocating authorship credit](#)” for the specific formulas). Two versions of the credit scores were calculated for the harmonic, geometric, and arithmetic allocation model:

- The standard version: the formulas for each model listed in Section “[Models for allocating authorship credit](#)” were used to calculate each author’s credit score as usual.
- The special version: the formulas for each model listed in Section “[Models for allocating authorship credit](#)” were first used to calculate each author’s credit score. The credit was then restructured between each paper’s authors so that the first and last authors of each paper share the credit for the first and second position, and the credit of the intermediate authors is reduced by one position (similar to what Hagen (2008, 2014a, 2014b) suggest). This version was calculated to accommodate the FLAE approach, and is considered a common practice in biomedicine and the lab-based life sciences (Larivière et al., 2016; Tschardt et al., 2007; Yank & Rennie, 1999).

¹ The taxonomy used for the final classification was tested for reliability and repeatability by enlisting an assisting coder to classify a subset of the data independently. The subset was chosen through disproportional stratified random sampling (10 of each heading) and was then structured in random order. The intercoder reliability was calculated as the number of agreements between the primary coder and the assisting coder divided by the number of agreements plus the number of disagreements. It reflects the percentage of data units on which coders agree.

Table 1 A three-tiered taxonomy for the classification of tasks found in the author contribution statement (ACS)

Type of task	Number of authors performing the task	Example text from the ACS of articles in Nature Chemical Biology
<i>Core layer task</i>	941	
Conception and design	749	"Designed the biological experiments"
Drafting article	616	"Wrote the manuscript with input from all of the authors"
<i>Middle layer task</i>	1695	
Assisting in the study design*	54	"Contributed to study design"
Assisting in drafting article*	401	"Contributed to editing the manuscript"
Collecting data	49	"Collected (X.AS data)"
Conducting experiments	1144	"Performed enzyme activity assays"
Data analysis	842	"Analyzed cellular growth assays"
<i>Outer layer task</i>	1058	
Advice, expertise or discussion	262	"Discussion of the chemical work"
Financial support	9	"Funded part of the research"
Supervision	134	"Supervised the collection and analysis of CD spectra"
Technical assistance	113	"Provided technical help"
Creating specimens/samples*	333	"Generated mutant receptors and cell lines"
Reviewing/proofing manuscript	272	"Reviewed the manuscript"
Other*	35	—
Programming	18	"Wrote scripts for automated photoswitching and imaging"
Project management	78	"Managed the overall project"
Providing equipment/software	11	"Provided the protocol for Mdm2 expression and purification"
Providing specimens/samples*	94	"Provided plasmids and cell lines"

The example text from the ACS and the number of authors performing a type of task are taken from the analyzed data set

*These types of tasks are modifications based on the preliminary analysis and are not found in the taxonomy used by Danell (2014)

The heading "Other" groups various contributions that occurred only sporadically in the data (nine times or less). Therefore, no example text representing this group is given since the group deals with many different but rare contributions

As the fractional and harmonic parabolic allocation models both give equal credit to the first and last author, there was no need to calculate a special version for them.

Relationship between empirical observations and author credit scores

To answer the second research question, I first constructed three definitions of an empirical observation:

- Definition A: author I's share of the total number of core-layer tasks specified in the author contribution statement of the paper.
- Definition B: author I's share of the total number of core- and middle-layer tasks specified in the author contribution statement of the paper.
- Definition C: author I's share of the total number of all tasks specified in the author contribution statement of the paper.

Standardized lack of fit was calculated for each model's standard and special version for all definitions in order to measure how well the models distribute credit to the authors. I also plotted the relationship between the empirical observations (using definition C) and the author's credit scores as calculated by the special version of the allocation models. The coefficients of determination (R^2) were calculated to measure how well each plotted allocation model replicates the empirical observations.

Receiver operating characteristic (ROC) analysis

The Receiver Operating Characteristic (ROC) framework is useful for evaluating the predictive value of bibliometric information (Lindahl & Danell, 2016; Zhang et al., 2019). Thus, in order to evaluate the different allocation models' performance in predicting the core authors (the third research question), an ROC analysis was performed in SPSS. The foundation of a ROC analysis is the confusion matrix illustrated in Table 2.

This matrix is a cross-tabulation of the predicted class versus the correct class of the instances in a data set (Fawcett, 2006; Provost & Fawcett, 2001). In this study, the correct class of each instance—or authorship, if you will—was derived from the analysis of author contributions statements, as described in Section “[Creating a data set](#)”. Each of the 1743 authorships in the data set was thus either classed (correctly) as being a core-layer author (true) or as not being a core-layer author (false).

The predicted class of each instance derives from the values of the different models for allocating authorship credit described in Section “[Models for allocating authorship credit](#)”. The credit scores produced for each author were considered the continuous output of a classifier that estimates membership in the core author group. A good classifier should give higher credit scores to core authors than to those given to middle- and outer-layer authors. Thus a higher score should indicate a higher probability of being a core author. As Fawcett (2006) describes, given a threshold parameter T , a predicted classification for an instance—positive or negative—can be extracted from such a classifier. A positive classification means the prediction is a core-layer author, and a negative means the prediction is a non-core-layer author (i.e., middle- or outer-layer author).

The two-by-two confusion matrix (exemplified in Table 2) classifies the authorships into four categories. Authorships predicted as positive can be classified as true positives (TP) if they belong to the core layer, or false positives (FP) if they do not belong to the core layer.

Table 2 The confusion matrix

	Correct class (i.e., being part of the core author layer)	
	True	False
Predicted class		
Positive	TP	FP
Negative	FN	TN

TP true positive, *FP* false positive, *FN* false negative, *TN* true negative

Similarly, authorships predicted as negative can be classified as false negatives (FN) if they belong to the core layer, or true negatives (TN) if they do not belong to the core layer.

From the confusion matrix, it is possible to calculate several common metrics. Below are two which are essential for understanding ROC:

$$\text{False Positive rate (fp rate)} = \frac{FP}{FP + TN}$$

$$\text{True Positive Rate (tp rate)} = \frac{TP}{TP + FN}$$

The tp and fp rates are strict columnar ratios and produce scores in the range [0, 1]. In ROC-space, which is defined by the tp rate on the y-axis and the fp rate on the x-axis, a single point is given by the values for the confusion matrix produced under the threshold T. For each classifier, a curve can be drawn in ROC-space by plotting the tp rate versus the fp rate for every possible value of T (starting with $+\infty$ and reducing step by step). Such a curve, called an ROC graph, depicts the relative tradeoffs between the benefits (true positives) and costs (false positives) of each classifier and has the attractive property of being “insensitive to changes in class distribution” (Fawcett, 2006, p. 864). The area under the ROC curve (AUC) is used to quantify each classifier’s performance in predicting the core authors. As stated in Fawcett (2006, p. 868), “the AUC of a classifier is equivalent to the probability that the classifier will rank a randomly chosen positive instance higher than a randomly chosen negative instance,” and it is used to compare the classifiers. In this study, the AUC for each author credit model can be restated as equivalent to the probability that a randomly chosen core-layer author receives a higher author credit score than a randomly chosen non-core-layer author.

Results

This section is structured into two parts. In the first part, I explore the relationship between an author’s position in the bylines of a research article and the contributions they have made. This is done to uncover what ordering logic structures articles in Chemical Biology (research question 1). In the second part, I test the validity of the five bibliometric counting methods (research questions 2 and 3).

The relationship between authorship order and research contributions

Table 3 shows that as the number of authors of each paper rises, the share of core authors positioned between the first and last author increases. This increase is because there only exist two unique author positions, the first and last position, while all author teams in the data set, without exception, have two or more core authors. When the number of authors on a paper reaches ten, the percentage of core authors in the middle position stops rising and seems to level out.

Regardless of author team size, almost all middle-layer authors are positioned between the first and last positions. Outer-layer authors have been left out of Table 3 as none of them figure as first or last authors. All authors performing only outer-layer tasks are found somewhere in the middle of the bylines (for more details, see Table 7).

As evidenced in Table 3, core authors do not only populate the first and last author positions. In addition, they are commonly found to inhabit the middle position, prompting the question of where the core authors that do not end up first or last in the bylines are positioned. Before answering this question, it is instructive to regard the average number of tasks performed as distributed per author position. Table 4 shows the average number of tasks for each author position as the number of total authors per publication (i.e., the author team) increases from 2 to 17. A striking feature of Table 4 is that first authors, regardless of the size of the author team, always perform more tasks on average than those in any other author position. Depending on the size of the author team, authors positioned last and authors positioned second are most often registered as the second- or third-highest performers of tasks on average. There are, however, two exceptions to this. First, when the author team is 14, the second-highest performer is found just before the last author position. Second, when the author team is 15, the second- and third-highest performer is found to be the second and third author position.

The general trend for each value of the author team is that the average number of tasks performed has a roughly u-shaped distribution. The average number of tasks in Table 4 decreases from the first author position until somewhere in the middle of the bylines, where it begins to increase and does so until it reaches the last author position. The reader can quickly see from the coloring in Table 4 that the position of the reds (high values) are at the beginning and the end of the bylines, and the blues (low values) are more in the middle of the bylines. However, the last author position never reaches the average number of tasks performed by the first authors, which again points to the importance of the first author in terms of workload and involvement in the research. The point where the decrease reaches the lowest value and turns into an increase differs between the sizes of the author team. For three authors, it is the second position that naturally has the lowest value, but for many other sizes of the author team, several positions report the lowest value. Sometimes the lowest average values of tasks performed correspond to the median position, but more times than not, the low point is slightly to the right of the median position. This result indicates that the u-shaped distribution is not symmetric.

Table 5 gives more detail on the specific positions of core-layer authors. Using the same coloring as in Table 4, we see that Table 5 reports a similar pattern; red (high values) on the two sides, followed by whiter shades (medium values) inwards the middle, and lastly, blue shades (low values) around the median position. This result tells us that the distribution of core-layer authors to specific positions in the byline is u-shaped and that this shape is a recurrent phenomenon regardless of the size of the author team.

Table 3 The percentages of first authors, last authors, and middle authors for core- and middle-layer authors

Authors of paper	Papers	Total authors	Core-layer authors			Middle-layer authors		
			First	Middle	Last	First	Middle	Last
2	14	28	50.0%	0.0%	50.0%	0.0%	0.0%	0.0%
3	18	54	40.9%	18.2%	40.9%	0.0%	100.0%	0.0%
4	15	60	32.5%	32.5%	35.0%	13.3%	80.0%	6.7%
5	10	50	30.8%	30.8%	38.5%	11.1%	88.9%	0.0%
6	23	138	24.4%	50.0%	25.6%	4.7%	93.0%	2.3%
7	20	140	24.7%	49.4%	26.0%	2.0%	98.0%	0.0%
8	13	104	23.5%	51.0%	25.5%	2.9%	97.1%	0.0%
9	14	126	23.1%	50.0%	26.9%	3.4%	96.6%	0.0%
10	17	170	14.9%	67.0%	18.1%	5.4%	94.6%	0.0%
11	13	143	16.7%	61.7%	21.7%	4.9%	95.1%	0.0%
12	7	84	19.4%	58.1%	22.6%	2.2%	97.8%	0.0%
13	13	169	15.9%	65.2%	18.8%	2.6%	97.4%	0.0%
14	8	112	16.3%	65.1%	18.6%	1.8%	98.2%	0.0%
15	9	135	13.6%	71.2%	15.3%	1.6%	98.4%	0.0%
16	8	128	17.6%	58.8%	23.5%	2.6%	97.4%	0.0%
17	6	102	16.7%	63.3%	20.0%	1.5%	98.5%	0.0%

First First author position, *Last* Last author position, *Middle* Middle author position

In other words, core-layer authors are in the majority in the positions that begin and end the bylines, while they are in the minority in the positions surrounding the median. The few positions that do not report any core authors are all median positions or positions close to the median.

In almost all sizes of author team, the position of last author is entirely made up of core-layer authors; The table indicates 100 percent for all author teams except those numbering 4 and 6. Comparing the first author position with the last author position, we see that the percentage of core-layer authors is higher in the latter group. This result is valid for all sizes of the author team except for sizes below four, where the percentages are equal. Furthermore, the percentages of core authors in each position are often higher on the right side of the median than on the left side. The conclusion is that the u-shape is nonsymmetric and that core-layer authors are more prevalent at the end of the bylines than at the beginning.

Table 6 gives more detail on the specific positions of middle-layer authors. The coloring is in stark contrast to Table 5, which reports the percentages of core authors; Instead of red (high values) on the edges, we have blue (low values), and instead of blue (low values) in the middle, we have red (high values). There is a gradual increase from the edges to the middle, where the highest reported value, more often than not, is to be found in the median position. Only for two sizes of the author team is the highest reported value not in, or directly next to, the median position. Instead of a u-shaped distribution, the distribution in Table 6 roughly resembles a bell-shaped distribution. There is no middle-layer author positioned last in the bylines (as we have seen from Table 5, core-layer authors populate the last author position). However, a small number populate the first author position. Middle-layer authors are more often positioned before the median than after, as the red in the table shows, pointing to a positive skewness in the distribution.

Table 4 The average number of tasks performed by an author on a specific author position

Authors of paper	Position of author																
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
2	4.5	3.3															
3	3.9	2.2	3.2														
4	4.3	3.3	2.4	3.3													
5	4.6	2.5	1.9	1.9	3.8												
6	4.7	3.2	2.3	2.1	2.4	3.7											
7	5.4	3.6	2.6	2.5	2.2	2.7	3.2										
8	4.3	3.8	2.2	1.7	1.7	1.5	2.2	3.2									
9	4.2	3.4	2.4	2.1	1.9	1.9	2.4	3.0	4.0								
10	4.4	3.5	3.0	2.1	1.8	1.5	1.9	2.4	2.8	3.4							
11	5.6	4.5	2.8	2.3	2.3	2.1	2.3	2.0	2.5	3.2	4.1						
12	4.6	4.4	2.9	2.1	2.1	2.1	2.0	1.7	1.7	2.4	3.3	3.6					
13	4.5	3.2	2.8	2.0	1.7	2.0	2.2	1.9	2.2	2.3	1.8	2.8	3.5				
14	5.3	2.8	2.6	2.6	2.5	2.8	1.9	1.8	2.0	2.4	2.8	3.0	3.8	3.5			
15	5.6	4.6	3.7	2.7	2.2	2.6	2.4	2.6	2.0	2.1	2.1	2.3	2.0	3.3	3.6		
16	5.9	4.0	2.1	3.0	3.3	2.0	2.0	2.3	1.9	2.6	2.6	2.6	1.9	2.1	3.4	3.8	
17	3.5	3.0	2.7	2.5	2.5	2.0	2.0	2.0	1.7	1.7	1.7	1.7	1.7	2.0	2.3	2.5	3.0

Note. The coloring goes from red (highest value) to white (50 percentile) to blue (lowest value).

Note. The boxes indicate the median position.

In Table 7, which gives more detail to the specific positions of outer-layer authors, there is a distinct lack of red (high values) and a prevalence of blue (low values). The values in Table 7 should be interpreted cautiously as the number of outer-layer authors is low; they only make up 11.0% of the whole data set. With that in mind, no outer-layer author is to be found first or last in the bylines. They are all spread out in the middle positions. For author team sizes up to 7, the median position reports the highest value, but it is hard to discern a pattern for the placement of outer-layer authors after that.

Any allocation model that does not account for the u-shaped distribution of tasks and core authors presented in this section is theoretically ill suited for allocating author credit in fields such as chemical biology. While fractional, harmonic, geometric, and arithmetic allocation does not do this, the harmonic parabolic allocation divides credit in this way and should theoretically be the best candidate to use. In the next section, the different allocation models are tested against how they perform in awarding higher credit to (1) the authors who make the greatest contributions in terms of the number of tasks they perform; and (2) the authors who are part of the core layer relative authors who are part of the middle and outer layer (i.e., how the models perform in predicting core authors).

Table 5 The percentage of core-layer authors in a specific author position

Authors of paper	Position of author																
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
2	100%	100%															
3	100%	44%	100%														
4	87%	47%	40%	93%													
5	80%	30%	20%	30%	100%												
6	91%	52%	30%	35%	70%	96%											
7	95%	55%	15%	15%	35%	70%	100%										
8	92%	69%	23%	8%	15%	23%	62%	100%									
9	86%	50%	14%	7%	14%	7%	14%	79%	100%								
10	82%	65%	41%	24%	29%	29%	47%	59%	76%	100%							
11	77%	62%	31%	15%	15%	23%	23%	23%	23%	69%	100%						
12	86%	43%	43%	0%	0%	0%	14%	0%	14%	57%	86%	100%					
13	85%	62%	31%	8%	8%	15%	23%	31%	38%	38%	23%	69%	100%				
14	88%	38%	25%	25%	25%	0%	13%	0%	25%	25%	38%	38%	100%	100%			
15	89%	67%	44%	22%	11%	22%	33%	22%	22%	33%	22%	44%	44%	78%	100%		
16	75%	50%	13%	13%	13%	13%	0%	0%	13%	13%	13%	13%	25%	13%	63%	100%	
17	83%	33%	17%	17%	17%	33%	17%	17%	17%	17%	17%	17%	17%	17%	33%	33%	100%

Note. The coloring goes from red (100%) to white (50%) to blue (0%).

Note. The boxes indicate the median position.

The validity of five bibliometric counting methods

Table 8 shows the standardized lack of fit between the authorship credit scores predicted by the five models and the share of tasks performed by each author according to the author contribution statement. There are three levels in the table: the first counts only core-layer tasks as the empirical observations; the second counts both core- and middle-layer tasks as the empirical observations; and the third counts all tasks as the empirical observations.

The first thing to notice in the table is that the special version of the arithmetic, geometric and harmonic models produces lower values of lack of fit at all levels of analysis. This finding is a direct result of the u-shaped distribution of tasks discussed earlier (see Table 4). Not taking this distribution into account makes for an author-credit distribution that underestimates the contributions of the last author. Three models—fractional, harmonic (special version), and harmonic parabolic—produce low values at all levels of counting, indicating a better fit between model predictions and empirical observations.

The second thing to notice in the table is the general decrease in lack of fit as we move from a more conservative to a more liberal criterion for defining empirical observations. As we move from counting only core-layer tasks to counting core- and middle-layer tasks and finally all types of tasks, empirical observations with a value of

Table 6 The percentages of middle-layer authors in a specific author position

Authors of paper	Position of author																
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
2	0%	0%															
3	0%	39%	0%														
4	13%	47%	33%	7%													
5	20%	50%	60%	50%	0%												
6	9%	43%	52%	57%	22%	4%											
7	5%	30%	70%	70%	50%	30%	0%										
8	8%	23%	69%	54%	62%	31%	15%	0%									
9	14%	36%	57%	93%	64%	79%	64%	14%	0%								
10	18%	29%	47%	59%	53%	59%	35%	18%	12%	0%							
11	23%	31%	38%	69%	62%	54%	62%	54%	54%	23%	0%						
12	14%	57%	29%	86%	86%	100%	86%	71%	71%	43%	0%	0%					
13	15%	38%	69%	69%	69%	54%	77%	54%	38%	54%	38%	15%	0%				
14	13%	50%	75%	63%	63%	100%	75%	88%	38%	63%	38%	50%	0%	0%			
15	11%	33%	56%	78%	67%	67%	44%	56%	56%	44%	56%	56%	44%	11%	0%		
16	25%	50%	75%	88%	88%	75%	100%	100%	63%	75%	50%	63%	38%	63%	25%	0%	
17	17%	50%	83%	83%	83%	67%	67%	83%	83%	67%	83%	67%	50%	67%	67%	67%	0%

Note. The coloring goes from red (100%) to white (50%) to blue (0%).

Note. The boxes indicate the median position.

zero decrease in number. None of the allocation models ever computes a zero value, which explains the general decrease in lack of fit.

Compared with the other models, the arithmetic allocation model never produces the highest or lowest values for lack of fit. The fractional model produces the lowest values of all models when counting all types of tasks, and it works well when counting both core- and middle-layer tasks. It also produces low values when counting core-layer tasks. The geometric model is consequently the worst in terms of lack of fit. As it is constructed to produce very low author credit scores for the authors not positioned at the beginning of the bylines, it means that even in its special version, it severely underestimates the contributions of many authors. At the same time, the model overestimates the relative contributions of authors positioned first. In its special version, the harmonic model performs better than the fractional model when counting core tasks. It also works well for the other two levels of analysis. The harmonic parabolic model produces the lowest values of all models both when counting core- and middle-layer tasks and when only counting core-layer tasks. It also produces low scores when counting all tasks.

Figure 1 plots the relationship between the empirical observations (i.e., author I's share of the total number of all tasks registered in the author contribution statement of the paper) and the author's credit scores, together with the corresponding coefficient of determination (R^2) for each allocation model. The special version of each model

Table 7 The percentages of outer-layer authors on a specific author position

Authors of paper	Position of author																
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
2	0%	0%															
3	0%	17%	0%														
4	0%	7%	27%	0%													
5	0%	20%	20%	20%	0%												
6	0%	4%	17%	9%	9%	0%											
7	0%	15%	15%	15%	15%	0%	0%										
8	0%	8%	8%	38%	23%	46%	23%	0%									
9	0%	14%	29%	0%	21%	14%	21%	7%	0%								
10	0%	6%	12%	18%	18%	12%	18%	24%	12%	0%							
11	0%	8%	31%	15%	23%	23%	15%	23%	23%	8%	0%						
12	0%	0%	29%	14%	14%	0%	0%	29%	14%	0%	14%	0%					
13	0%	0%	0%	23%	23%	31%	0%	15%	23%	8%	38%	15%	0%				
14	0%	13%	0%	13%	13%	0%	13%	13%	38%	13%	25%	13%	0%	0%			
15	0%	0%	0%	0%	22%	11%	22%	22%	22%	22%	22%	0%	11%	11%	0%		
16	0%	0%	13%	0%	0%	13%	0%	0%	25%	13%	38%	25%	38%	25%	13%	0%	
17	0%	17%	0%	0%	0%	0%	17%	0%	0%	17%	0%	17%	33%	17%	0%	0%	0%

Note. The coloring goes from red (100%) to white (50%) to blue (0%).
 Note. The boxes indicate the median position.

Table 8 Standardized lack of fit between the authorship credit scores predicted by the five models and the proportion of tasks performed by each author according to the author contribution statement

	Arithmetic	Fractional	Geometric	Harmonic	Harmonic parabolic
Counting only core-layer tasks					
Standard version	0.969687	0.212305	207.3667	0.517496	0.123427
Special version	0.308563	0.212305	40.28017	0.16846	0.123427
Counting only core- and middle-layer tasks					
Standard version	0.152211	0.047994	20.00023	0.094942	0.046433
Special version	0.081224	0.047994	9.198605	0.057543	0.046433
Counting all tasks					
Standard version	0.110808	0.025627	17.19636	0.07204	0.033281
Special version	0.064957	0.025627	10.78752	0.046907	0.033281

Standard version. The credit scores for each allocation model have been calculated according to the formulas listed in Section “Models for allocating authorship credit”

Special version. The credit scores for each allocation model have been calculated according to the formulas listed in Section “Models for allocating authorship credit” but with the additional rule that the first and last author receives equal credit scores

is used as it outperforms the standard version in all cases according to Table 8. As reported in Fig. 1, the geometric model performs worst in replicating the empirical observations ($R^2 = 0.4494$). The harmonic ($R^2 = 0.6168$), fractional ($R^2 = 0.6326$), and

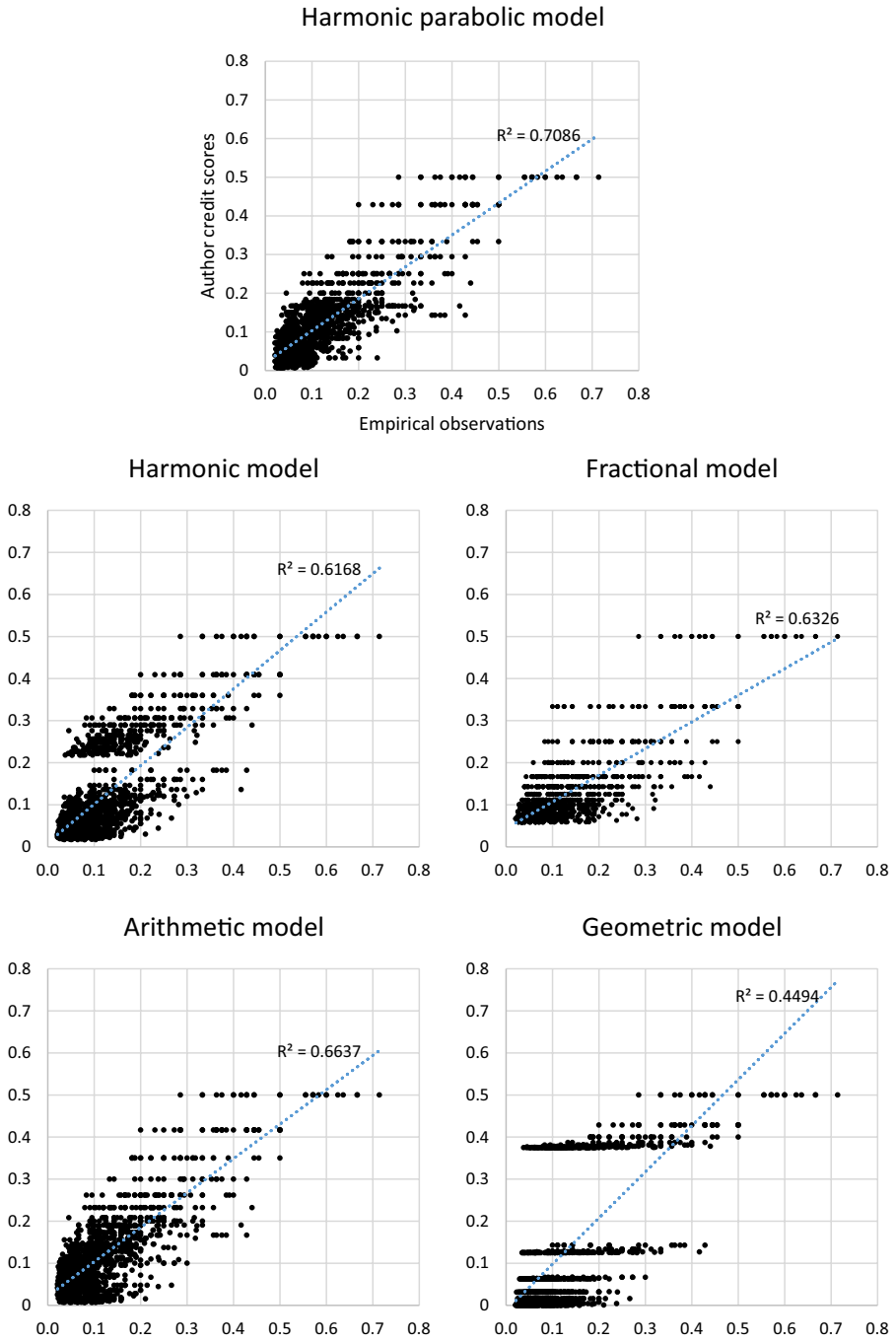


Fig. 1 Relationship between empirical observations and author credit scores produced by five counting models. The empirical observations are defined as author I's share of the total number of all tasks registered in the author contribution statement of the paper. The author credit scores are calculated so that the first and last author receives equal credit scores

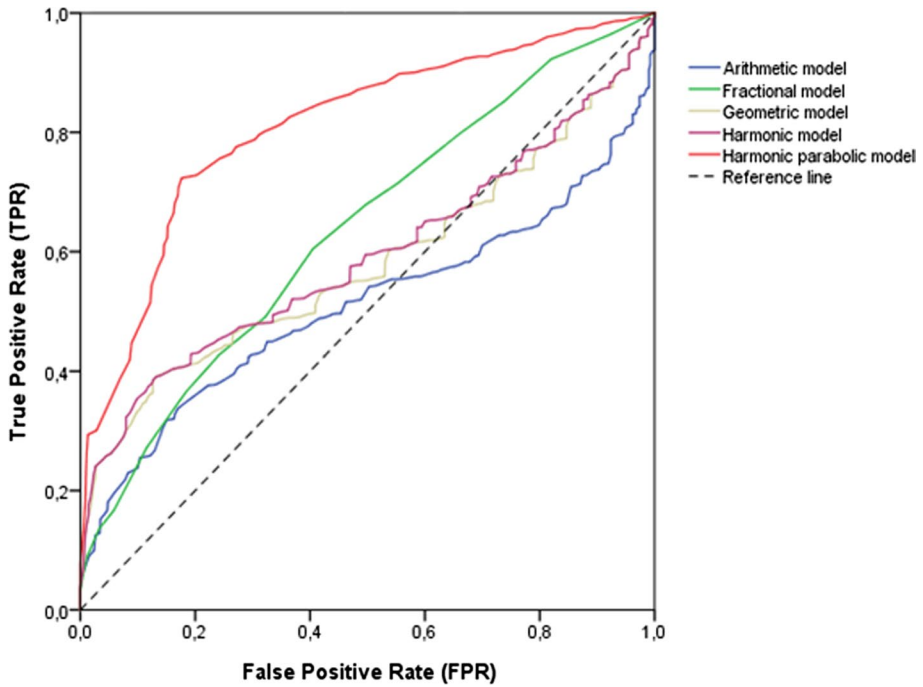


Fig. 2 ROC curve for predicting core authors (normal version). The credit scores for each allocation model have been calculated according to the formulas listed in Section "Models for allocating authorship credit"

Table 9 Area Under the Curve (for ROC curve in Fig. 2)

Models for allocating authorship credit	AUC	Std. error ^a	Asymptotic Sig. ^b	Asymptotic 95% Confidence Interval	
				Lower bound	Upper bound
Arithmetic (Standard)	0.509	0.015	0.533	0.480	0.537
Fractional	0.635	0.013	0.000	0.609	0.661
Geometric (Standard)	0.580	0.014	0.000	0.552	0.608
Harmonic (Standard)	0.591	0.014	0.000	0.564	0.619
Harmonic Parabolic	0.810	0.010	0.000	0.789	0.830

The test result variables: Fractional, Geometric, Arithmetic, Harmonic, and Harmonic Parabolic, have at least one tie between the positive actual state group and the negative actual state group. Statistics may be biased

^aUnder the nonparametric assumption

^bNull hypothesis: true area = 0.5

arithmetic model ($R^2 = 0.6637$) all give similar values, while the harmonic parabolic model ($R^2 = 0.7086$) does the best job of explaining the variation in the empirical data.

Figures 2 and 3 show the ROC curve indicating how well the different allocation models predict the core authors. Figure 2 uses the author credit scores produced by the standard version, and Fig. 3 uses the author credit scores for the special version, which

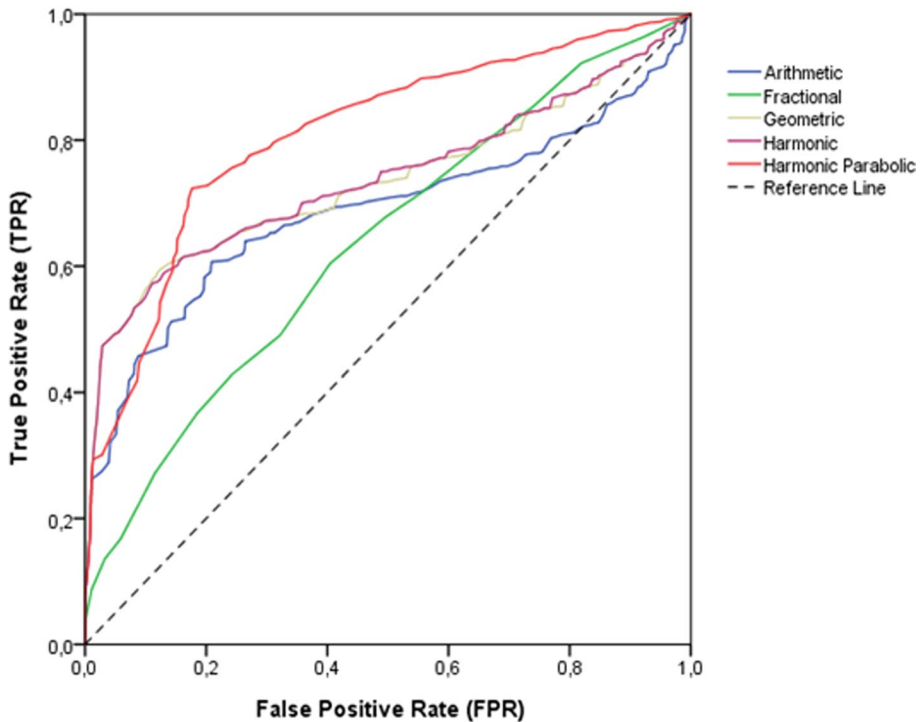


Fig. 3 ROC curve for predicting core authors (special version). The credit scores for each allocation model have been calculated according to the formulas listed in Section "Models for allocating authorship credit" but with the additional rule that the first and last author receives equal credit scores

Table 10. Area Under the Curve (for ROC curve in Fig. 3)

Models for allocating authorship credit	AUC	Std. error ^a	Asymptotic Sig. ^b	Asymptotic 95% Confidence Interval	
				Lower bound	Upper Bound
Arithmetic (Special)	0.686	0.014	0.000	0.659	0.712
Fractional	0.635	0.013	0.000	0.609	0.661
Geometric (Special)	0.735	0.013	0.000	0.710	0.760
Harmonic (Special)	0.738	0.013	0.000	0.714	0.763
Harmonic Parabolic	0.810	0.010	0.000	0.789	0.830

The test result variable(s): Arithmetic, Fractional, Geometric, Harmonic, and Harmonic Parabolic have at least one tie between the positive actual state group and the negative actual state group. Statistics may be biased

^aUnder the nonparametric assumption

^bNull hypothesis: true area = 0.5

treats the first and last authors' credit as equal. Tables 9 and 10 give the area under the curve (AUC), which measures how well the models differentiate between core- and non-core authors based on their credit scores. As in Table 8, when considering the lack

of fit between author credit scores and percentages of tasks performed by an author, the special version produces ROC curves with a larger area under the curve. For the standard version (Fig. 2, Table 9), no model other than the harmonic parabolic does a compelling job at predicting which authors are core authors and which are not. The fractional model performs better than the other three but still performs poorly. In the special version (Fig. 3, Table 10) of the models, the geometric and the harmonic do a fair job of separating core- and non-core authors. However, the harmonic parabolic model still performs best. Specifically, when using the Harmonic Parabolic model, the probability that a randomly chosen core-layer author receives a higher author credit score than a randomly chosen non-core-layer author is 81%.

Discussion and conclusions

Scientific authorship serves as a means of establishing priority, getting peer acknowledgment (Merton, 1973), and assigning accountability for scientific truth claims (Biagioli & Galison, 2003). Due to the increasing number of authors per article (Larivière et al., 2015), the increasing diversity in the meaning ascribed to authorship (Smith et al., 2019, 2020), and the perceived increase in scientific fraud and irreproducible research (Begley & Ioannidis, 2015; Steen, 2011), alternative ways of attributing scientific authorship has become an ongoing topic for discussion by stakeholders in science (Larivière et al., 2021). One suggestion from these discussions is that authors disclose their scientific contributions in an article's author contribution statement (Rennie et al., 1997). In the present paper, such author contribution statements form the basis for an analysis of the author-ordering logic in the field of chemical biology, as well as an analysis of the validity of five bibliometric counting methods in this field.

The ordering of authors in Chemical Biology implies differences in relative contributions. When ordering the authorships according to the position on the bylines, there is a distinct u-shaped distribution both for (a) the percentage of authors involved in writing the paper or designing the research (i.e., core authors), and (b) the average number of tasks performed. All sizes of author team report this. The two distributions have some differences, however. While the percentages of core authors are higher at the end than at the beginning of the bylines, the inverse is detected for the average number of tasks. Borrowing the terminology of Baerlocher et al. (2007), it thus seems that the analyzed data uncovers two important but different types of authors: *primary authors* positioned at the beginning of the bylines (e.g., the first and second author), who do more of the research than the authors equivalently positioned at the end of the bylines (e.g., the last and second from last author); and *supervisory authors* positioned at the end of the bylines, who are more often involved in core tasks such as writing the manuscript and research design than the authors equivalently positioned at the beginning of the bylines. The authors around the median position are seldom core authors. Such positions also report low values for the average number of tasks performed. If one aims to allocate credit by the importance of an author's contributions in the field of Chemical Biology, the formula for calculating such credit must consider the above u-shaped distributions. It should give less credit to authors in the median position while increasing the credit given as one approaches the first and last positions, which is in line with the author-ordering logic for this field.

It is probable that the distributions detailed above are related to different demographic variables such as gender, academic age, and rank. Larivière et al. (2016) report that

contributions that I classify as part of the core-layer are associated with high mean academic age. In contrast, middle-layer tasks such as analyzing data and performing experiments are associated with a lower mean year from first publication (Larivière et al., 2016). According to Costas & Bordons (2011) there is a strong trend for younger researchers and researchers in the lower academic ranks to figure as first authors, while senior researchers and researchers with high academic status are more likely to be positioned last. Larivière et al. (2021) can identify a gendered divide between conceptual and empirical work. For example, conceptualizing a study, something I classify as a core-layer task, is more likely to be performed by men. Women are more likely to contribute to investigation (i.e. the middle-layer task of performing the experiments). While interesting, such factors were not part of the data for this paper and thus could not be analyzed. However, I aim to explore them using a larger and more diverse dataset in a future study.

Standardized lack of fit was calculated for each model's standard and special version in order to measure how well the author-credit allocation models distribute credit to the authors. As is to be expected, there is no difference between the values reported by the standard and the special version of the fractional and the harmonic parabolic models. However, calculating the arithmetic, geometric and harmonic model as normal, as opposed to calculating the credit for the first and last authors as equal in each model, produces significantly worse results. This is a direct result of the u-shaped distribution of tasks. When using definition C (i.e., author I's share of the total number of all tasks registered in the author contribution statement of the paper) for the empirical observations, the fractional model does the best job distributing author credit. However, for both definition B (i.e., author I's share of the total number of core- and middle-layer tasks registered in the author contribution statement of the paper) and A (i.e., author I's share of the total number of core-layer tasks registered in the author contribution statement of the paper), the harmonic parabolic model performs best. The harmonic parabolic model also performs better than the special version of the other models in replicating the empirical observations (definition C), as measured by the coefficient of determination. Theoretically, it would seem that using only core-layer tasks or core- and middle-layer tasks as the empirical observation does a better job at defining an important author in terms of contributions. This is because the core- and middle-layer tasks are similar to those used by scientific organizations such as the International Committee of Medical Journal Editors (ICMJE) to determine which authors are worthy of inclusion in the bylines. Counting an author's share of the total number of tasks in the author contribution statement introduces several outer-layer tasks of lesser importance for attributing authorship (Biagioli & Galison, 2003; Danell, 2014; House & Seeman, 2010; Sundling, 2017). It also gives them equal value to the core- and middle-layer tasks. However, it does not rely on classifying tasks into different layers, so it is much less subject to classification bias (e.g., Lambert, 2011).

Calculating the author allocation models according to the special version (i.e., equating the credit of the first and last author) as opposed to the standard version produces a significant increase in the area under the curve for the arithmetic, geometric and harmonic models. Regardless, the arithmetic model (together with the fractional model) performs poorly, while the geometric and harmonic models do a fair job of separating core- and non-core authors. However, the harmonic parabolic model outperforms them all, doing a good job of predicting which authors are core authors and which are not.

I would argue that the harmonic parabolic model is the one that performs best overall for the analyzed data set, even though the fractional model produces the lowest lack of fit when considering all types of tasks. Looking at the reported coefficients of determination in Fig. 1 and the ROC curves in Figs. 2 and 3, together with the area under the curve in

Tables 9 and 10, lends proof to the consideration that the harmonic parabolic model is the best overall pick for allocating author credit in bibliometric exercises in the field of chemical biology.

There are some limitations of this study. The most critical problem is accurately measuring the extent of a contribution, which is something that even a detailed author contribution statement has problems expressing. I have made efforts to classify contributions as core-, middle-, or outer-layer, which partly reflects the extent of involvement in a task. However, it is important to realize that even though an author could be involved in more tasks than another author, it is not sure that the former has contributed more than the latter. Furthermore, even when contributors are classified as performing the same type of task, there is a risk that there is a considerable difference in their extent of involvement. Such misclassifications ultimately depend on how specific the authors have been when detailing the contributions and if the classification procedure picks this up. On an aggregate level, the errors introduced by these should be lower, which motivates further studies using larger datasets—something I aim to take on in the near future. However, looking at the data for this study, core-layer authors are, on average, involved in almost twice as many tasks as middle-layer authors and dominate the first and last author positions. So even though uncertainties exist concerning the extent of involvement, the u-shaped author distributions align with what we know about the organization of lab work and author ordering (e.g. Larivière et al., 2016; Louis et al., 2008; Sundling, 2017).

Another limitation is that this study is based only on articles in the chemical biology field; as such, the results apply only within the context of this field. As detailed investigations into all byline positions are lacking in the literature, there is no room for saying that the harmonic parabolic model performs as well in other parts of science as it does in chemical biology. However, there are indications that papers in biomedicine and clinical medicine also have several primary and supervisory authors positioned at the beginning and the end of the bylines, respectively (Mongeon et al., 2017). Baerlocher et al. (2007) and Wren et al. (2007) make similar conclusions in these fields; They found that first authors contributed the most, followed by last and second authors, while authors situated in the middle had low levels of participation in tasks such as conception and drafting the manuscript. Recent studies of author contribution statements from articles in several PLOS journals also suggest that first and last authors contribute to more tasks than middle authors (Larivière et al., 2021; Sauermann & Haeussler, 2017). In fields with such an ordering logic, the harmonic parabolic model would probably perform better than an allocation model that assumes a declining ordering logic, such as the harmonic model.

Author inclusion and byline order vary per field, but it is crucial to acknowledge that these aspects also likely vary between labs, specializations, and countries (Knorr-Cetina, 1999; Pontille, 2003; Smith et al., 2019, 2020; Trimbur & Braun, 1992). For example, some lab leaders might be more inclined to “rank an unlucky post-doc first on a paper on which he or she would normally have been placed only as a second author, in order to keep that person motivated” (Knorr-Cetina, 1999, p. 231). Furthermore, there are indications that transdisciplinary research collaborations have more inclusive authorship practices than research where the collaborators are from the same field (Elliott et al., 2017). Such factors also influence the findings and generalizability of this study.

This study evaluates the validity of different models for allocating authorship credit by comparing them in their standard form, and when considering the first and last positions equal. One might argue that treating the credit for corresponding authors as equal to that of the first and last authors would have produced other results and corrected some models.

There are, however, reasons for not including such a version of the models in the analysis. First, while it is true that many articles in the data set contain multiple corresponding authors, only one for each paper is registered in the reprint author field in WOS (Waltman, 2012). As long as this is the case, including all corresponding authors for each paper in an allocation model to be used in more extensive bibliometric studies would be a somewhat unrealistic enterprise. Second, the reprint authors that WOS reports for this data set all appear as either first or last authors, making a different version of the models redundant.

Many studies have shown that not fractionalizing authorship credit produces inflated numbers which fundamentally change rankings (Gauffriau & Larsen, 2005; Gauffriau et al., 2008; Huang et al., 2011; Piro et al., 2013). Hagen (2014a) points out that while the effects of this kind of inflationary bias are well known, the same cannot be said for the effect of distributing credit equally among coauthors who have not contributed equally (i.e., the equalizing bias). Some studies indicate that using a harmonic allocation model, as opposed to a fractional model, produces rankings and metrics that are different on the level of individual researchers (Hagen, 2014a) as well as for countries (Hagen, 2015). A future research venue lies in studying rank and output metrics changes when the harmonic parabolic model is used instead of the harmonic or the fractional model. This line of inquiry could also be combined with data about the gender of the authors as well as academic rank and organizational prestige to uncover distributional biases in the author credit allocation models. Another fruitful enterprise would be using author contribution data from journals adopting the CRediT-taxonomy for a large-scale study of allocation models in various fields.

Appendix

Example of an author contribution statement

“D. Sakano performed chemical screening, cellular and biochemical analyses; D. Sakano, N.S. and K.U. established the ES cell differentiation system; D. Sakano, K. Kikawa, M.K. and T.Y. performed transplantation assays; K.A. established the ES cell line; S.M., F.E. and N.N. helped maintain AKITA mice; D.M. and M.U. provided and analyzed the chemical library; O.A. and D. Stainier provided chemicals; K. Kume and S.K. provided technical advices, S.K. designed the experiments and wrote the paper. All of the authors discussed the results and commented on the manuscript.” (Sakano et al., 2014).

Acknowledgements I like to acknowledge Drs. Richard Danell and Cristian Colliander for comments on this manuscript. A big thanks also goes out to the reviewers whose comments greatly improved the paper.

Funding Open access funding provided by Umea University. The author has no financial or personal relationship with other people or organizations that could inappropriately influence or bias the paper's content. No grants or funding (save for a salary paid by the Department of Sociology) has been received for this work.

Declarations

Conflict of interest The paper contains original research and is submitted with full responsibility: There is no fraud or plagiarism. It has not been published earlier in any journal, and it is not being considered for publication elsewhere.

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

- Abbott, A., Cyranoski, D., Jones, N., Maher, B., Schiermeier, Q., & Van Noorden, R. (2010). Metrics: Do metrics matter? *Nature News*, *465*(7300), 860–862. <https://doi.org/10.1038/465860a>
- Akhabue, E., & Lautenbach, E. (2010). “Equal” contributions and credit: An emerging trend in the characterization of authorship. *Annals of Epidemiology*, *20*(11), 868–871. <https://doi.org/10.1016/j.annepidem.2010.08.004>
- Assimakis, N., & Adam, M. (2010). A new author's productivity index: P-index. *Scientometrics*, *85*(2), 415–427. <https://doi.org/10.1007/s11192-010-0255-z>
- “Authorship : authors & referees @ npg.” (2015). Authorship: Authors & referees @ npg. Retrieved December 10, 2022, from <http://www.nature.com/authors/policies/authorship.html>
- Aziz, N. A., & Roziq, M. P. (2013). Profit (p)-index: The degree to which authors profit from co-authors. *Plos One*, *8*(4), e59814. <https://doi.org/10.1371/journal.pone.0059814>
- Baerlocher, M. O., Newton, M., Gautam, T., Tomlinson, G., & Detsky, A. S. (2007). The meaning of author order in medical research. *Journal of Investigative Medicine*, *55*(4), 174–180. <https://doi.org/10.2310/6650.2007.06044>
- Begley, C. G., & Ioannidis, J. P. (2015). Reproducibility in science: Improving the standard for basic and preclinical research. *Circulation Research*, *116*(1), 116–126.
- Biagioli, M., & Galison, P. (2003). *Scientific authorship: Credit and intellectual property in science*. Routledge.
- Birnholtz, J. P. (2006). What does it mean to be an author? The intersection of credit, contribution, and collaboration in science. *Journal of the American Society for Information Science and Technology*, *57*(13), 1758–1770. <https://doi.org/10.1002/asi.20380>
- Costas, R., & Bordons, M. (2011). Do age and professional rank influence the order of authorship in scientific publications? Some evidence from a micro-level perspective. *Scientometrics*, *88*(1), 145–161. <https://doi.org/10.1007/s11192-011-0368-z>
- 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>
- Danell, R. (2014). The contribution of authors: A study of the relationship between the size and composition of author teams. *Proceedings of the Science and Technology Indicators Conference 2014, 3 - 5 September, Leiden, the Netherlands*, 123–131. Retrieved December 10, 2022, from <http://umu.diva-portal.org/smash/record.jsf?pid=diva2%3A770323&dsid=5190>
- Davenport, E., & Cronin, B. (2001). Who dunnit? Metatags and hyperauthorship. *Journal of the American Society for Information Science and Technology*, *52*(9), 770–773. <https://doi.org/10.1002/asi.1123>
- Egghe, L., Rousseau, R., & Van Hooydonk, G. (2000). Methods for accrediting publications to authors or countries: Consequences for evaluation studies. *Journal of the American Society for Information Science*, *51*(2), 145–157. [https://doi.org/10.1002/\(SICI\)1097-4571\(2000\)51:2%3c145::AID-ASI6%3e3.0.CO;2-9](https://doi.org/10.1002/(SICI)1097-4571(2000)51:2%3c145::AID-ASI6%3e3.0.CO;2-9)
- Elliott, K. C., Settles, I. H., Montgomery, G. M., Brassel, S. T., Cheruvelil, K. S., & Soranno, P. A. (2017). Honorary authorship practices in environmental science teams: Structural and cultural factors and solutions. *Accountability in Research*, *24*(2), 80–98. <https://doi.org/10.1080/08989621.2016.1251320>
- Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, *27*(8), 861–874. <https://doi.org/10.1016/j.patrec.2005.10.010>
- Frandsen, T. F., & Nicolaisen, J. (2010). What is in a name? Credit assignment practices in different disciplines. *Journal of Informetrics*, *4*(4), 608–617. <https://doi.org/10.1016/j.joi.2010.06.010>

- Gaufrriau, M., & Larsen, P. O. (2005). Counting methods are decisive for rankings based on publication and citation studies. *Scientometrics*, 64(1), 85–93. <https://doi.org/10.1007/s11192-005-0239-6>
- Gaufrriau, M., Larsen, P. O., Maye, I., Roulin-Perriard, A., & von Ins, M. (2008). Comparisons of results of publication counting using different methods. *Scientometrics*, 77(1), 147–176. <https://doi.org/10.1007/s11192-007-1934-2>
- Hagen, N. T. (2008). Harmonic allocation of authorship credit: Source-level correction of bibliometric bias assures accurate publication and citation analysis. *PLOS ONE*, 3(12), e4021. <https://doi.org/10.1371/journal.pone.0004021>
- Hagen, N. T. (2010). Harmonic publication and citation counting: Sharing authorship credit equitably—not equally, geometrically or arithmetically. *Scientometrics*, 84(3), 785–793. <https://doi.org/10.1007/s11192-009-0129-4>
- Hagen, N. T. (2013). Harmonic coauthor credit: A parsimonious quantification of the byline hierarchy. *Journal of Informetrics*, 7(4), 784–791. <https://doi.org/10.1016/j.joi.2013.06.005>
- Hagen, N. T. (2014a). Counting and comparing publication output with and without equalizing and inflationary bias. *Journal of Informetrics*, 8(2), 310–317. <https://doi.org/10.1016/j.joi.2014.01.003>
- Hagen, N. T. (2014b). Reversing the byline hierarchy: The effect of equalizing bias on the accreditation of primary, secondary and senior authors. *Journal of Informetrics*, 8(3), 618–627. <https://doi.org/10.1016/j.joi.2014.05.003>
- Hagen, N. T. (2015). Contributory inequality alters assessment of academic output gap between comparable countries. *Journal of Informetrics*, 9(3), 629–641. <https://doi.org/10.1016/j.joi.2015.06.002>
- Hicks, D. (2012). Performance-based university research funding systems. *Research Policy*, 41(2), 251–261. <https://doi.org/10.1016/j.respol.2011.09.007>
- Hodge, S. E., & Greenberg, D. A. (1981). Publication credit. *Science*, 213(4511), 950–950. <https://doi.org/10.1126/science.213.4511.950.b>
- Hosseini, M., & Bruton, S. V. (2020). A qualitative study of equal co-first authorship. *Accountability in Research*, 27(8), 496–520. <https://doi.org/10.1080/08989621.2020.1776122>
- House, M. C., & Seeman, J. I. (2010). Credit and authorship practices: Educational and environmental influences. *Accountability in Research: Policies & Quality Assurance*, 17(5), 223–256. <https://doi.org/10.1080/08989621.2010.512857>
- Hu, X. (2009). Loads of special authorship functions: Linear growth in the percentage of “equal first authors” and corresponding authors. *Journal of the American Society for Information Science and Technology*, 60(11), 2378–2381. <https://doi.org/10.1002/asi.21164>
- Huang, M.-H., Lin, C.-S., & Chen, D.-Z. (2011). Counting methods, country rank changes, and counting inflation in the assessment of national research productivity and impact. *Journal of the American Society for Information Science and Technology*, 62(12), 2427–2436. <https://doi.org/10.1002/asi.21625>
- Jian, D., & Xiaoli, T. (2013). Perceptions of author order versus contribution among researchers with different professional ranks and the potential of harmonic counts for encouraging ethical co-authorship practices. *Scientometrics*, 96(1), 277–295. <https://doi.org/10.1007/s11192-012-0905-4>
- Knorr-Cetina, K. (1999). *Epistemic cultures: How the sciences make knowledge*. Harvard University Press.
- Lambert, J. (2011). Statistics in brief: How to assess bias in clinical studies? *Clinical Orthopaedics and Related Research*, 469(6), 1794–1796. <https://doi.org/10.1007/s11999-010-1538-7>
- 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., Gingras, Y., Sugimoto, C. R., & Tsou, A. (2015). Team size matters: Collaboration and scientific impact since 1900. *Journal of the Association for Information Science and Technology*, 66(7), 1323–1332. <https://doi.org/10.1002/asi.23266>
- Larivière, V., Pontille, D., & Sugimoto, C. R. (2021). Investigating the division of scientific labor using the contributor roles taxonomy (CRediT). *Quantitative Science Studies*, 2(1), 111–128. https://doi.org/10.1162/qss_a_00097
- Lindahl, J., & Danell, R. (2016). The information value of early career productivity in mathematics: A ROC analysis of prediction errors in bibliometrically informed decision making. *Scientometrics*. <https://doi.org/10.1007/s11192-016-2097-9>
- Liu, X. Z., & Fang, H. (2012). Fairly sharing the credit of multi-authored papers and its application in the modification of h-index and g-index. *Scientometrics*, 91(1), 37–49. <https://doi.org/10.1007/s11192-011-0571-y>
- Louis, K. S., Holdsworth, J. M., Anderson, M. S., & Campbell, E. G. (2008). Everyday ethics in research: Translating authorship guidelines into practice in the bench sciences. *The Journal of Higher Education*, 79(1), 88–112.

- Lukovits, I., & Vinkler, P. (1995). Correct credit distribution: A model for sharing credit among coauthors. *Social Indicators Research*, 36(1), 91–98. <https://doi.org/10.1007/BF01079398>
- Merton, R. K. (1973). *The sociology of science: Theoretical and empirical investigations*. University of Chicago Press.
- Mongeon, P., Smith, E., Joyal, B., & Larivière, V. (2017). The rise of the middle author: Investigating collaboration and division of labor in biomedical research using partial alphabetical authorship. *PLOS ONE*, 12(9), e184601. <https://doi.org/10.1371/journal.pone.0184601>
- Morris, S. A., & Goldstein, M. L. (2007). Manifestation of research teams in journal literature: A growth model of papers, authors, collaboration, coauthorship, weak ties, and Lotka's law. *Journal of the American Society for Information Science and Technology*, 58(12), 1764–1782. <https://doi.org/10.1002/asi.20661>
- Ostler, E. L. (2007). Chemical biology is.... *Chemistry Central Journal*. <https://doi.org/10.1186/1752-153X-1-5>
- Piro, F. N., Aksnes, D. W., & Rørstad, K. (2013). A macro analysis of productivity differences across fields: Challenges in the measurement of scientific publishing. *Journal of the American Society for Information Science and Technology*, 64(2), 307–320. <https://doi.org/10.1002/asi.22746>
- Pontille, D. (2003). Authorship practices and institutional contexts in sociology: Elements for a comparison of the United States and France. *Science, Technology, & Human Values*, 28(2), 217–243. <https://doi.org/10.1177/0162243902250905>
- Price, D. D. S. (1981). Multiple authorship. *Science*, 212(4498), 986. <https://doi.org/10.1126/science.212.4498.986-a>
- Provost, F., & Fawcett, T. (2001). Robust classification for imprecise environments. *Machine Learning*, 42(3), 203–231. <https://doi.org/10.1023/A:1007601015854>
- Rennie, D., Yank, V., & Emanuel, L. (1997). When authorship fails: A proposal to make contributors accountable. *JAMA*, 278(7), 579–585. <https://doi.org/10.1001/jama.1997.03550070071041>
- Sakano, D., Shiraki, N., Kikawa, K., Yamazoe, T., Kataoka, M., Umeda, K., Araki, K., Mao, D., Matsumoto, S., Nakagata, N., Andersson, O., Stainier, D., Endo, F., Kume, K., Uesugi, M., & Kume, S. (2014). VMAT2 identified as a regulator of late-stage β -cell differentiation. *Nature Chemical Biology*, 10(2), 141–148. <https://doi.org/10.1038/nchembio.1410>
- Sauermann, H., & Haeussler, C. (2017). Authorship and contribution disclosures. *Science Advances*, 3(11), e1700404. <https://doi.org/10.1126/sciadv.1700404>
- Shapiro, D., Wenger, N., & Shapiro, M. (1994). The contributions of authors to multiauthored biomedical research papers. *JAMA*, 271(6), 438–442. <https://doi.org/10.1001/jama.1994.03510300044036>
- Smith, E., Williams-Jones, B., Master, Z., Larivière, V., Sugimoto, C. R., Paul-Hus, A., Shi, M., Diller, E., Caudle, K., & Resnik, D. B. (2020). Researchers' perceptions of ethical authorship distribution in collaborative research teams. *Science and Engineering Ethics*, 26(4), 1995–2022. <https://doi.org/10.1007/s11948-019-00113-3>
- Smith, E., Williams-Jones, B., Master, Z., Larivière, V., Sugimoto, C. R., Paul-Hus, A., Shi, M., & Resnik, D. B. (2019). Misconduct and misbehavior related to authorship disagreements in collaborative science. *Science and Engineering Ethics*, 26(4), 1967–1993. <https://doi.org/10.1007/s11948-019-00112-4>
- Stallings, J., Vance, E., Yang, J., Vannier, M. W., Liang, J., Pang, L., Dai, L., Ye, I., & Wang, G. (2013). Determining scientific impact using a collaboration index. *Proceedings of the National Academy of Sciences*, 110(24), 9680–9685. <https://doi.org/10.1073/pnas.1220184110>
- Steen, R. G. (2011). Retractions in the scientific literature: Is the incidence of research fraud increasing? *Journal of Medical Ethics: Journal of the Institute of Medical Ethics*, 37, 249–253. <https://doi.org/10.1136/jme.2010.040923>
- Sundling, P. (2017). The many hands of science: Commonalities and differences in the research contributions of authors and subauthors. *Aslib Journal of Information Management*, 69(5), 591–606. <https://doi.org/10.1108/AJIM-01-2017-0012>
- Trimbur, J., & Braun, L. A. (1992). Laboratory life and the determination of authorship. *New visions of collaborative writing* (pp. 19–36). Boynton/Cook Publishers.
- Trueba, F. J., & Guerrero, H. (2004). A robust formula to credit authors for their publications. *Scientometrics*, 60(2), 181–204. <https://doi.org/10.1023/B:SCIE.0000027792.09362.3f>
- Tscharntke, T., Hochberg, M. E., Rand, T. A., Resh, V. H., & Krauss, J. (2007). Author sequence and credit for contributions in multiauthored publications. *PLoS Biol*, 5(1), e18. <https://doi.org/10.1371/journal.pbio.0050018>
- Van Hooydonk, G. (1997). Fractional counting of multiauthored publications: Consequences for the impact of authors. *Journal of the American Society for Information Science*, 48(10), 944–945. [https://doi.org/10.1002/\(SICI\)1097-4571\(199710\)48:10%3C944::AID-ASIS%3e3.0.CO;2-1](https://doi.org/10.1002/(SICI)1097-4571(199710)48:10%3C944::AID-ASIS%3e3.0.CO;2-1)

- Vavryčuk, V. (2018). Fair ranking of researchers and research teams. *PLOS ONE*, 13(4), e0195509. <https://doi.org/10.1371/journal.pone.0195509>
- Waltman, L. (2012). An empirical analysis of the use of alphabetical authorship in scientific publishing. *Journal of Informetrics*, 6(4), 700–711. <https://doi.org/10.1016/j.joi.2012.07.008>
- Waltman, L. (2016). A review of the literature on citation impact indicators. *Journal of Informetrics*, 10(2), 365–391. <https://doi.org/10.1016/j.joi.2016.02.007>
- Wilsdon, J., Allen, L., Belfiore, E., Campbell, P., Curry, S., Hill, S., Jones, R., Kain, R., Kerridge, S., Thelwall, M., Tinkler, J., Viney, I., Wouters, P., Hill, J., & Johnson, B. (2015). *The metric tide: Independent review of the role of metrics in research assessment and management*. Sage.
- Wren, J. D., Kozak, K. Z., Johnson, K. R., Deakyne, S. J., Schilling, L. M., & Dellavalle, R. P. (2007). The write position. *EMBO Reports*, 8(11), 988–991. <https://doi.org/10.1038/sj.embor.7401095>
- Wuchty, S., Jones, B. F., & Uzzi, B. (2007). The increasing dominance of teams in production of knowledge. *Science*, 316(5827), 1036–1039. <https://doi.org/10.1126/science.1136099>
- Xu, J., Ding, Y., Song, M., & Chambers, T. (2016). Author credit-assignment schemas: A comparison and analysis. *Journal of the Association for Information Science and Technology*, 67(8), 1973–1989. <https://doi.org/10.1002/asi.23495>
- Yang, S., Wolfram, D., & Wang, F. (2017). The relationship between the author byline and contribution lists: A comparison of three general medical journals. *Scientometrics*, 110(3), 1273–1296. <https://doi.org/10.1007/s11192-016-2239-0>
- Yank, V., & Rennie, D. (1999). Disclosure of researcher contributions: A study of original research articles in the Lancet. *Annals of Internal Medicine*, 130(8), 661–670.
- Zhang, Y., Wang, M., Gottwalt, F., Saberi, M., & Chang, E. (2019). Ranking scientific articles based on bibliometric networks with a weighting scheme. *Journal of Informetrics*, 13(2), 616–634. <https://doi.org/10.1016/j.joi.2019.03.013>