Investigation of the external validity of the 2004 German Science Foundation author contribution calculation recommendation for medical schools' performance-based funding systems

Paul Donner*

*donner@dzhw.eu
ORCID: 0000-0001-5737-8483
Department 2 'Research System and Science Dynamics', German Centre for Higher Education Research and Science Studies (DZHW), Germany

This study examines how well an idiosyncratic authorship counting rule for co-authored publications recommended by the German Science Foundation (DFG) for medical schools and widely used in performance-based funding systems aligns with the empirical evidence. The DFG rule and two other co-author credit rules are compared with empirical data of percentage contribution statements of authors of co-authored papers in medicine.

1. Introduction

For the past decades, bibliometrically informed performance-based funding systems (PBFS) have been established and became entrenched at medical schools in Germany. These systems are used to partially allocate shares of block funding from the federal states internally to subunits of a medical school. As there is no central evaluation authority, medical schools have been free to design and implement systems according to their own preferences for criteria, weights, rules, and so on, so that there is now considerable diversity of systems across schools. Practically all of them involve the counting of publications or impact factor points of publications in some way, alongside other criteria, with the intention of rewarding and encouraging internationally competitive academic research.

In 2004 the Senate Commission for Clinical Research of the German Science Foundation (Deutsche Forschungsgemeinschaft, DFG), the largest and most science-politically influential research funding organization in Germany, published a recommendation for the design of PBFS at German medical schools (DFG, 2004) which would prove quite consequential. In the following years, many medical schools adjusted their systems to align with the DFG recommendations and newly established systems also often followed them.

The DFG made several recommendations, including the use of 3-year running averages for calculations, distributing 20–40 % of the block funding competitively, using obtained third-party funding and publications as criteria, discouraging criteria which in their opinion do not directly reflect a unit's own research performance (completed PhDs and habilitations, patents, offices held, book chapters, editorships, and others). Concerning publications, the recommendations criticized the exclusive use of Journal Impact Factors (p. 14) and instead promoted the development of adequate methods to judge the scientific quality of individual publications. However, it left open the back door by allowing to use Impact Factors in a 'surrogate manner' in the 'intervening time' until such methods have been developed.

Furthermore, with the stated intention to prevent excessive multiple counting of the same paper by several subunits by simply including more co-authors to manipulate such a procedure, it was recommended to count first authors with a weight of one third of the Impact Factor score, last authors also with one third, and to split the remaining third among all other 'middle' authors (p. 15). This somewhat idiosycratic rule was not justified by any reference to published

evidence about the contribution of authors of various author positions in medicine. It was also not recommended to use this method for anything else than internal PBFS calculations. The method is quite similar to that of Abramo, Cicero, and D'Angelo (2015), used for the life sciences, in which first and last authors get equal credit, while middle authors split the remainder, the exact values depending furthermore on whether a publication is an intramural or extramural collaboration.

The DFG's proposed counting rule recommendation has found relatively broad acceptance in Germany. Aman and Van den Besselaar (2024) recently surveyed German medical schools regarding their PBFS rules. They found that, of 18 responding schools, 9 followed the DFG rule exactly while 7 others followed it with some adaptation.

One might expect that any rule that divided the credit of a publication among co-authors should align as well as possible with how much authors in different author positions typically contribute to a paper. In this contribution we will therefore analyze if this particular counting method, which we will henceforth refer to as the "DFG rule of thirds", is supported by the available empirical data of relative shares of co-author contributions in medicine. For comparison, we will confront alternative co-author counting methods which were designed more or less explicitly to reflect authors' relative contributions, with the same data.

2. Data and methods

We use micro-level data on the estimated explicitly and publicly stated relative contributions of authors to co-authored publications from Donner (2020). These percentage contribution data were collected from the full texts of cumulative PhD theses – theses made up of several individually published papers. Some university departments require explicit contribution statements for co-authored papers in theses in order for PhD committees to assess the overall contribution of the PhD candidate. Such statements, checked, accepted and signed by all relevant co-authors and therefore considered highly reliable, were searched. Data were collected for 124 theses from 22 universities in three countries. Theses were categorized by discipline based on the thesis title and PhD granting department. Contribution data for 93 authorships in 72 publications related to 19 medical dissertations were found and only these will be analyzed here. Fig. 1 shows the data points by author position.

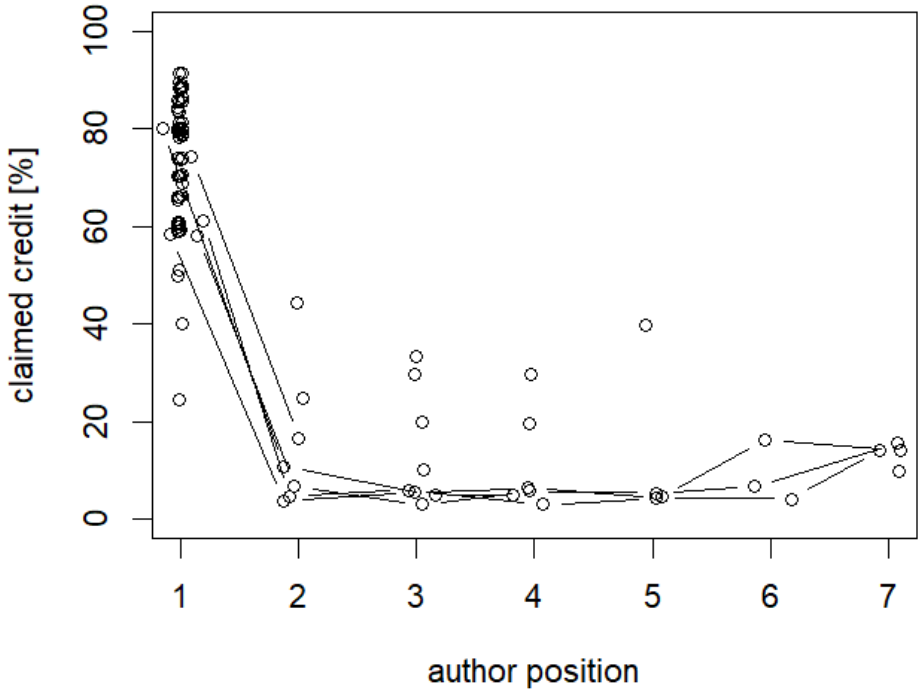
A co-author credit counting method should produce results that are as close as possible to real co-author contributions. We test for this using correlation coefficients and calculating the average absolute errors between counting method values and observed empirical values. That means we use the empirical data as the external criterion to establish the criterion validity of counting methods as measurement methods of relative co-author contributions. The absolute error between two values, say a contribution of 28 percent and a counting method's value of 40 percent would be: $|28 - 40| = 12$. We report the average of these errors as percentage points.

Although the DFG recommendation does not mention the case of papers with two authors, we assign both authors a contribution share of 50 % in those papers. For comparison purposes with the rule of thirds, we test the counting rule currently most commonly used in bibliometrics, fractional counting, which simply assigns the same value of $1/n$ to each co-author, where n is the number of co-authors. We further test the validity of harmonic counting against the data, which was previously found to have the best fit to empirical data from several counting methods (Donner, 2020). Harmonic counting (Hagen, 2008; Hodge & Greenberg, 1981), assigns decreasing values of credit to greater author positions according to the formula

$$harmonic\ credit = \frac{\frac{1}{i}}{1+\frac{1}{2}+\cdots+\frac{1}{n}},$$

where *i* is the co-author position.

Figure 1: Empirical data of claimed credit in percent over author positions. Small random values added to make points distinguishable. Lines connect data points of one publication



3. Results

The data contain 61 observations of first authorships. These contributed on average 73 % to a paper (range: 25–90 %). There are only 6 observations for last authorships. These contributed on average 18 % (range: 5–33 %). While there were 26 observations for middle authors, it makes little sense to calculate an average, as these are publications with different numbers of middle authors. Clearly, the DFG rule of thirds overestimates the contribution of last authors while strongly underestimating that of first authors.

We now examine how well the rule of thirds and alternative counting methods correlate with the empirical data. We calculated correlation coefficients of the raw empirical micro-level data and for averages of contributions aggregated by author count and author position. The latter is done because the micro level data is not distributed equally across author positions. For example, for 4-author papers, there are six observations for first author position but only one each for second to fourth positions.

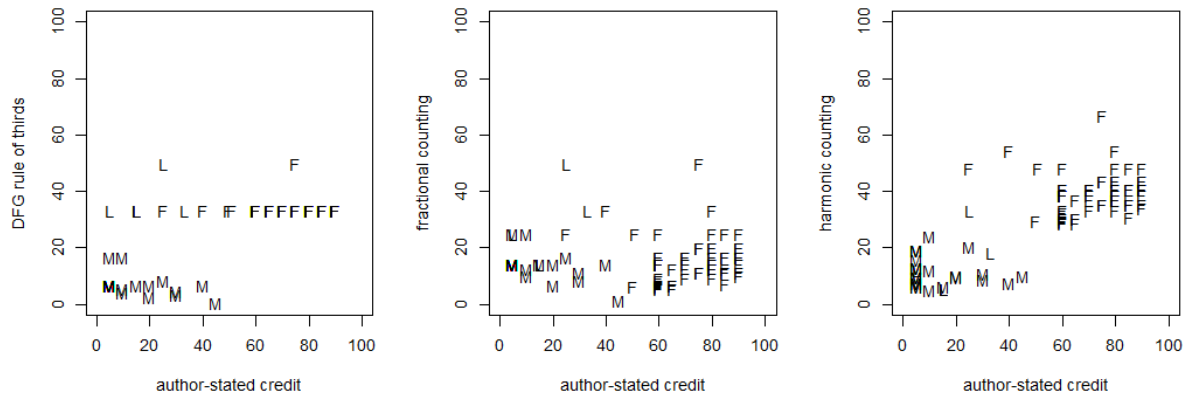
Table 1: Agreement of three co-author counting methods with empirical data from medicine

	micro-level data (N=93)		aggregated by author count and position (N=34)
counting method	**correlation coefficient**	**average absolute error [in percentage points]**	**correlation coefficient**
DFG rule of thirds	0.75 ($p < 0.01$)	29.6	0.66 ($p < 0.01$)
(equal) fractional counting	–0.02 ($p = 0.83$)	41.3	−0.09 ($p = 0.61$)
harmonic counting	0.82 ($p < 0.01$)	26.3	0.78 ($p < 0.01$)

The results are presented in Table 1. The correlation of the DFG rule of thirds with the data is high (.75, 95 % CI: .65-.83) but not as high as that of harmonic counting (.82, CI: .73-.87). Fractional counting is uncorrelated with authors' credit claims (r= –.02, CI: –.23-.18) and will misjudge the true contribution on average by about 41 percentage points. The rule of thirds will be off by about 30 points, harmonic counting will be off by about 26 points.

Fig. 2 shows the relationships between the values of three compared counting rules and the empirical data on co-author contributions. Author positions are coded as letters as follows: "F": first author, "M": middle author, "L": last author. It can be recognized quite clearly that both the rule of thirds and fractional counting fail to appropriately use the upper end of the range of contributions, thereby underestimating first authors' work shares.

Figure 2: Scatterplots of co-author counting rules values vs. empirical data (N=93)



4. Discussion and conclusion

Our results should be interpreted with some caution due to the limited size of the dataset, in particular, there are few observations for last authors. The data is also not a random sample and it is not representative across the actual author count and position distribution of medicine publications. Clearly, more contribution claims data should be collected and analyzed. Nevertheless, the evidence shows that the simple rule of thirds may be too simple, as it assigns too much weight to last authors at the expense of first authors, that is, it is apparently biased against first authors. Perhaps surprisingly, it still conforms far better to the data than the current standard method of professional bibliometrics, fractional counting. A more sophisticated counting rule giving different credit according to author position, harmonic counting, gives

results much more in line with the empirical data. Still, according to its correlation with the studied data, a better credit allocation method might exist or be developed.

On the basis of these findings we suggest that the DFG and German medical schools reconsider their recommendations and implemented co-author counting rules. Towards this end, dedicated studies collecting more empirical data of medical researchers' estimated percentage contributions to co-authored works should be carried out to have a more reliable evidence base for such consequential policies. The frequently used method of full counting (Aman and Van den Besselaar, 2024) is not a viable alternative because its relative distribution is the same as the fractional counting method studied here (equal values for all co-authors), which was found to be uncorrelated with authors' stated contributions.

Open science practices
The data used in this study is freely available at https://zenodo.org/records/3755227.

Competing interests
The author declares no competing interest.

Funding information
This research has not benefited from any external funding.

References
Abramo, G., Cicero, T., & D'Angelo, C. A. (2015). Should the research performance of scientists be distinguished by gender? *Journal of Informetrics*, *9*(1), 25-38.

Aman, V., & Van den Besselaar, P. (2024). Authorship regulations in performance-based funding systems and publication behaviour–A case study of German medical faculties. *Journal of Informetrics*, *18*(2), 101500.

DFG (2004). *Empfehlungen zu einer Leistungsorientierten Mittelvergabe (LOM) an den Medizinischen Fakultäten. Stellungnahme der Senatskommission für Klinische Forschung*, Bonn.

Donner, P. (2020). A validation of coauthorship credit models with empirical data from the contributions of PhD candidates. *Quantitative Science Studies*, *1*(2), 551-564.

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.

Hodge, S. E., & Greenberg, D. A. (1981). Publication credit. *Science*, *213*, 950.