To what extent does the citation advantage of collaboration depend on the citation counting system?

Jonathan M. Levitt[1], Mike Thelwall[2] and Michael Levitt[3]

[1] *j.levitt@lboro.ac.uk and j.levitt@wlv.ac.uk*
Department of Information Science, University of Loughborough, Ashby Road, Loughborough, Leicestershire LE11 3TU (UK) and
Statistical Cybermetrics Research Group,
University of Wolverhampton, Wulfruna Street, Wolverhampton WV1 1LY (UK)

[2] *m.thelwall@wlv.ac.uk*
School of Technology, University of Wolverhampton, Wulfruna Street, Wolverhampton WV1 1LY (UK)

[3] *michael.levitt@stanford.edu*
Department of Structural Biology, Stanford University School of Medicine, Stanford, CA 94305 (USA)

Abstract

This paper investigates the extent to which the apparently higher productivity and citation impact of authors with high levels of co-authorship depend on the way in which citation credit is assigned. For articles in physics it compares two methods of allocating credit, the standard (whole) counting system and the fractional counting system. Although numerous studies have found that when the whole counting system is used, more highly co-authoring researchers are more productive and publish more highly cited articles, no study has established whether this is true when the fractional counting system is used. This is particularly important as findings on co-authorship can be used to support the recent encouragement of co-authorship by research policy. This study finds: (a) when whole counting is used, the productivity and citation level increases with level of co-authorship; but (b) when fractional counting is used, the productivity and citation level decreases with level of co-authorship. This casts doubt on the received wisdom about the advantage of co-authored research.

Introduction

In recent years co-authorship has been encouraged by research policy and, over the last fifty years, research in sciences, social sciences, and the humanities has become increasingly co-authored (Wuchty, Jones & Uzzi, 2007). Associated with this support of co-authorship, bibliometric research has found that, in general, co-authored articles are more highly cited than solo authored articles. In addition, several researchers have examined different ways of assigning credit to co-authored articles. This paper assesses in novel ways the apparent bibliometric advantage of co-authorship and the allocation of credit to co-authors. It focuses on highly co-authored science, which is often the result of well funded large-scale research projects, and hence the findings are particularly relevant for funding bodies.

Related Research and Research Questions

Research evaluation allocates credit for co-authored research in three ways: (1) gives full credit to every author of the article; (2) shares a single credit for the article equally amongst all of the authors; (3) shares a single credit unequally amongst all the authors –typically allocating the most credit to the first author. In (2), each author of an article with 20 co-authors is assigned a twentieth of the credit of an article by a sole author; (3) is not examined in this paper, as it has been used less often than the other ways, and several variants have been suggested (Hagen, 2010). The first method, called ‘whole counting’ or ‘standard counting’, is widely used in both peer review and bibliometrics (e.g., the UK’s Research Excellence Framework, the h-index and many studies of co-authorship). The second method, called ‘fractional counting’, is used in some bibliometric investigations of co-authorship; it was recommended in the 1980s (e.g., Lindsey, 1980; Price, 1981) and has been used or

investigated in numerous articles (e.g., Burrell & Rousseau, 1995; Glänzel & De Lange, 2002; Gauffriau & Larsen, 2005; Sooryamoorthy, 2009). Our paper examines how the decision whether to use whole or fractional counts affects findings on the apparent advantage of co-authorship.

Several studies have found a positive relationship between level of co-authorship and productivity. More than eighty years ago, Lotka (1926) found that, in general, people in larger author partnerships published more. Lotka's findings were confirmed by subsequent studies (e.g., Price & Beaver, 1966; Zuckerman, 1967; Pao, 1982; Pravdic & Oluic-Vukovic, 1986; Lee & Bozema, 2005). Lee and Bozema (2005), however, in their survey of US scientists, checked how their findings would be affected by fractional counting. They found that, with this normalization, the number of authors is not a significant predictor of publishing productivity but did not specifically examine highly co-authored research and hence its advantage is still unclear. This gap is addressed by the following research question:

Question 1: Do researchers with high levels of co-authorship publish more articles than researchers with lower levels of co-authorship?

Numerous studies have found that articles by a larger number of authors are generally more highly cited than articles by fewer authors. For example, higher levels of citation have been found for Chilean physics (Vogel 1997), chemistry (Glänzel & Schubert 2001), Danish industry (Frederiksen 2004), library and information science (Levitt & Thelwall, 2009) and economics (Levitt & Thelwall, 2010). In their comparison of whole and fractional counting, Egghe, Rousseau and Van Hooydonk (2000) showed mathematically that different systems can yield very different rankings; however, neither they, nor any other study of co-authorship, have fully investigated the extent to which the higher citation of collaborative research depends on the counting system. In order to fill this gap, our paper investigates how fractional counting would affect the findings. Specifically, it examines:

Question 2: To what extent are researchers with high levels of co-authorship more highly cited than researchers with lower levels of co-authorship?

By investigating Questions 1 and 2 in two different counting systems this study seeks to answer the following question that is relevant to the decision by many policy makers to encourage co-authored research:

Question 3: To what extent does the apparent productivity and citation advantage of co-authored research depend on the counting system used?

Data and Methods

Our research investigates a subset of articles published in 2004 in the Science Citation Index (SCI), namely articles with at least one author's address from the UK in one or more of the following SCI subject categories: Astronomy & Astrophysics; Physics, Multidisciplinary; Physics, Particles & Fields; Physics, Nuclear. This subset consists of 3,364 articles and for brevity is called 'Physics_AstroNuclear'. Being restricted to a single productive country, the UK, makes the analysis small enough to be manageable while still large enough to be significant. It was decided to focus on Physics_AstroNuclear articles, as its component subjects were each found to have particularly many highly co-authored articles and thus enables the investigation of a wide range of levels of co-authorship and the inclusion of findings relevant to highly collaborative science.

The Physics_AstroNuclear set described above was identified as follows. First, using the Analyse facility of the Web of Science (WoS) the number of articles published by the 500th most prolific author was obtained for each of the 16 SCI subject categories that have the most articles published in 2004 with UK addresses. The 500th most prolific author published 25 articles in 'Astronomy & Astrophysics', whereas for not one of the other 15 subjects did the 500th most prolific author publish more than 3 articles in 2004. Next, on the basis of these

findings, the other eleven subjects containing the letters 'physic' were examined; 'physic' was chosen in order to identify not only subjects containing the word 'physics', but also subjects containing the word 'Physical'. The 500th most prolific author published 32 articles in 'Physics, Multidisciplinary', 25 articles in 'Physics, Particles & Fields' and 5 articles in 'Physics, Nuclear', whereas for not one of the other 8 subjects did the 500th most prolific author publish more than 2 articles in 2004.

The most prolific author in Physics_AstroNuclear published 82 articles in 2004, and the 500th most prolific author 56 articles, and 422 authors published exactly 57 articles in that year. The Appendix presents the distribution of all articles with over 50 authors and some data on the 234 articles with at least 50 authors and some data on the 183 articles with over 200 authors.

In order to obtain the data for Tables 1 and 2, the records of the 3,364 Physics_AstroNuclear articles were downloaded from the online interface of WoS and the authors and total citation of each article extracted from the records. The data was analysed by first calculating for each author the average level of authorship, and then evaluating the productivity and citation level for both whole and fractional counting. Then the averages of this data were calculated for all the authors in non-overlapping ranges of authorship, thereby enabling the relationships between authorship and productivity and between authorship and citation to be compared for the two counting systems.

Findings

Tables 1 and 2 contain the data used to address all three research questions. Both tables present findings on whole and fractional productivity ('Article credit per author') and citation level ('Citation credit per author') for diverse levels of co-authorship ('Author range').

Table 1: Productivity and citation data for all authors with an average number of co-authors of less than 20.

Co-author level	*Average level*	*Authors*	*Article credit per author*		*Citation credit per author*	
			Whole	*Fractional*	*Whole*	*Fractional*
1 - 2	1.1	181	1.6	1.40	9.5	8.76
2 - 3	2.1	765	1.6	.78	16.8	8.15
3 - 4	3.1	1,057	1.7	.55	19.8	6.51
4 - 5	4.1	936	1.6	.42	24.2	6.01
5 - 6	5.1	766	1.6	.35	29.1	5.88
6 - 7	6.1	566	1.6	.28	25.8	4.34
7 - 8	7.1	534	1.6	.26	24.7	3.61
8 - 9	8.1	362	1.5	.22	22.7	2.98
9 - 10	9.1	298	1.7	.21	28.7	3.30
10 - 11	10.1	294	1.7	.20	27.4	2.85
11 - 12	11.1	224	2.0	.23	26.2	2.50
12 - 13	12.2	172	2.0	.19	50.8	4.31
13 - 14	13.1	206	1.7	.16	28.9	2.45
14 - 15	14.1	174	1.8	.14	39.7	2.93
15 - 16	15.1	155	1.8	.15	26.9	1.89
16 - 17	16.1	100	2.1	.17	30.8	2.17
17 - 18	17.1	124	1.8	.15	29.9	1.85
18 - 19	18.1	119	2.0	.15	33.3	1.92
19 - 20	19.1	131	1.9	.13	30.0	1.68

In each table, 'Co-author level' denotes the range in which the number of co-authors falls. The value of co-authorship is greater than or equal to the lower value and less than the upper value; for example in Table 1, the row with Average co-authors '1 - 2' presents the findings on all authors that have an average level of co-authorship greater than or equal to 1 and less than 2. The difference between the two tables is that Table 1 presents detailed findings on the authors with an Author range of less than 20, whereas Table 2 presents broader findings on the authors with average co-authorship of less than 630; the 45 authors with average co-authorship of over 630 were excluded as because more than 80% (37) co-authored the same single article.

Table 2: Productivity and citation data for all authors with an average number of co-authors of less than 630.

Co-author level	*Average level*	*Authors*	*Article credit per author*		*Citation credit per author*	
			Whole	*Fractional*	*Whole*	*Fractional*
1 - 30	9.3	8,349	1.8	.36	26.4	4.49
30 - 60	40.9	1,033	2.0	.08	44.8	1.19
60 - 90	70.6	486	2.0	.05	125.0	1.85
90 - 120	101.4	422	2.4	.04	54.8	.57
120 - 150	132.2	544	3.5	.03	81.1	.61
150 - 180	160.8	93	2.8	.09	300.5	2.17
180 - 210	200.0	551	4.7	.03	85.1	.44
210 - 240	230.8	505	8.8	.06	47.9	.27
240 - 270	251.1	71	10.4	.11	22.1	.32
270 - 300	281.8	301	5.9	.03	21.8	.09
300 - 330	311.5	446	8.8	.05	35.0	.16
330 - 360	345.7	1,141	10.5	.04	99.5	.30
360 - 390	368.4	757	8.5	.02	73.7	.20
390 - 420	401.2	28	13.0	.06	49.6	.19
420 - 450	438.6	26	14.3	.06	36.9	.09
450 - 480	469.4	55	8.1	.06	26.0	.19
480 - 510	503.4	348	9.8	.04	42.7	.10
510 - 540	533.3	268	16.7	.04	33.6	.08
540 - 570	554.0	188	17.0	.04	33.2	.06
600 - 630	614.0	380	13.9	.07	47.4	.08

The fourth and fifth columns indicate productivity. 'Whole' is evaluated by adding the number of articles published by each author in the Author range and dividing by the number of authors in the Author range. 'Fractional' is evaluated by adding for every author the author's total fractional for publication count and dividing by the number of authors in the level of Co-authorship ('Co-author level'); 'Average level' is evaluated by adding the number of authors of the articles by each author in the Co-author level and dividing by the average number of authors in the Co-author level.

The sixth and seventh columns indicate citation level. 'Whole' is evaluated by adding the number of citations of the articles published by each author in the Author range and dividing by the number of articles in the Author range. 'Fractional' is evaluated by adding for every author the author's total fractional citation count and dividing by the number of authors in Co-author level.

We present an example that illustrates the whole and fractional values: Assume that an author publishes the following three articles: (a) article A is by one author and cited twice, (b) article B is by two authors and cited six times and (c) article C is by four authors and cited four

times. Then the author's whole article credit is 3 (number of articles), fractional article credit is 1.75 (1+.5+.25), whole citation credit is 12 (sum of citations) and fractional citation credit is 6 (2+3+1).

Question 1, the extent to which researchers with higher levels of co-authorship produce more articles, is addressed using the data in columns four and five. Whole productivity increases steadily with level of co-authorship. Fractional productivity decreases substantially when co-authorship ranges from 1 to 30 and undulates (between .02 and .11) in co-authorship ranges from 30 to 630.

Question 2, the extent to which researchers with higher levels of co-authorship produce more citations, is addressed using the data in columns six and seven. Whole citation increases steadily as co-authorship increases from 1 to 4, apart from one outlier undulates (between 22 and 40) as co-authorship increases from 5 to 20, and undulates (between 21 and 300) as co-authorship increases above 30. Fractional citation decreases steadily as co-authorship rises from 1 to 9, undulates (between .44 and 2.17) as co-authorship rises from 30 to 180, decreases as co-authorship rises from 180 to 270, and undulates (between .08 and .30) co-authorship increases above 270.

Discussion

The main limitations of this research are that the data are from a specific time period (2004) and a subset of science (astrophysics and related articles with at least one UK co-author). Whilst the subset topic seems to be appropriate for analysing highly co-authored articles, it may not be representative for co-authorship trends for low numbers of co-authors. Although this research is at the individual rather than institutional level, its findings are based on large sets of individuals, and so do not suffer from limitations often found in investigations at the individual level.

Question 3 asks to what extent the productivity and citation advantages of co-authored research depend on the way in which co-authorship is treated. The Findings indicate that both the productivity and citation advantages of co-authored research depend on the way in which co-authorship is counted.

Whole counting indicates: (a) productivity increases gradually as the number of authors increases; (b) level of citation increases substantially when the number of authors increases from 1 to 4. In contrast, fractional counting indicates: (a) productivity decreases substantially when co-authorship ranges from 1 to 60; (b) citation decreases when co-authorship ranges from 1 to 9. These findings indicate that the productivity and citation advantages or disadvantages of co-authorship depend critically on the counting system.

In order to compare the relative strengths of this contrasting behaviour, the geometric means of the whole and fractional article credit and citation credit from Tables 1 and 2 are presented in Figures 1 and 2. Figure 1 indicates that, apart from the decline in article fractional credit from 1 to 8 authors, the trends counteract each other.

The number of co-authors is a factor that influences not only the number of citations for one paper but also the number of papers per researcher: In general larger co-author groups publish more articles and receive more citations. Therefore, similar to field normalizing techniques, in principle citation analysis could normalise for the number of co-authors.

The finding that, when using the fractional counting system, productivity and citation impact decreases with the number of authors does not cast doubt on previous findings that collaborative articles are in general more highly cited. However it does provide a fuller picture of the relative merits of collaborative research: Collaborative articles generally are more highly cited, but the number of citations per author decreases with increased co-authorship. This latter finding raises concerns on the merits of encouraging collaboration. Specifically if, for a given subject and year of publication, *citations per researcher* indicate

value for money in research expenditure, then this research finds that greater value for money would be obtained from less collaborative research.

The assumption that, for a given subject and year of publication, the metric 'citations per researcher' indicates value for money in research could be regarded as simplistic. But if this assumption is made, then this study indicates that, in general, smaller groups of co-authors are *better* value for money in research than larger groups of co-authors. Naturally this finding needs to be treated with caution, as it is based on a simplistic assumption and on the investigation of a single subject. This study is not intended to discourage collaborative research, but to provide a fuller picture of the relative merits of co-authorship.

Conclusions

Studies of collaboration generally have used the whole counting system and, using this system, there is substantial evidence that more highly co-authored articles are more highly cited. This has reinforced the concept that collaborative research is more highly cited and by implication likely to be of a higher quality. However, when the fractional counting system is used, the citation of co-authors declines as the number of authors increases from 1 to 9. A corollary of this is that a scientist wishing to make the most publications or generate the most citation impact should stick to single authored research and should definitely not engage in highly co-authored research, where their individual contribution will generate, once the contributions of co-authors are accounted for, minimal outputs and minimal impact. This is a very abstract argument, however, and does not take into account the value of research to society. Moreover, each of the co-authors of a large research project might be essential to its smooth functioning.

In this context, Wuchty, Jones and Uzzi (2007) wrote "We have used 19.9 million papers over 5 decades and 2.1 million patents to demonstrate that teams increasingly dominate solo authors in the production of knowledge. Research is increasingly done in teams across nearly all fields." It would be ironic if this move towards increased co-authorship has been influenced by findings on co-authorship that reflect the way in which credit for research has been attributed.

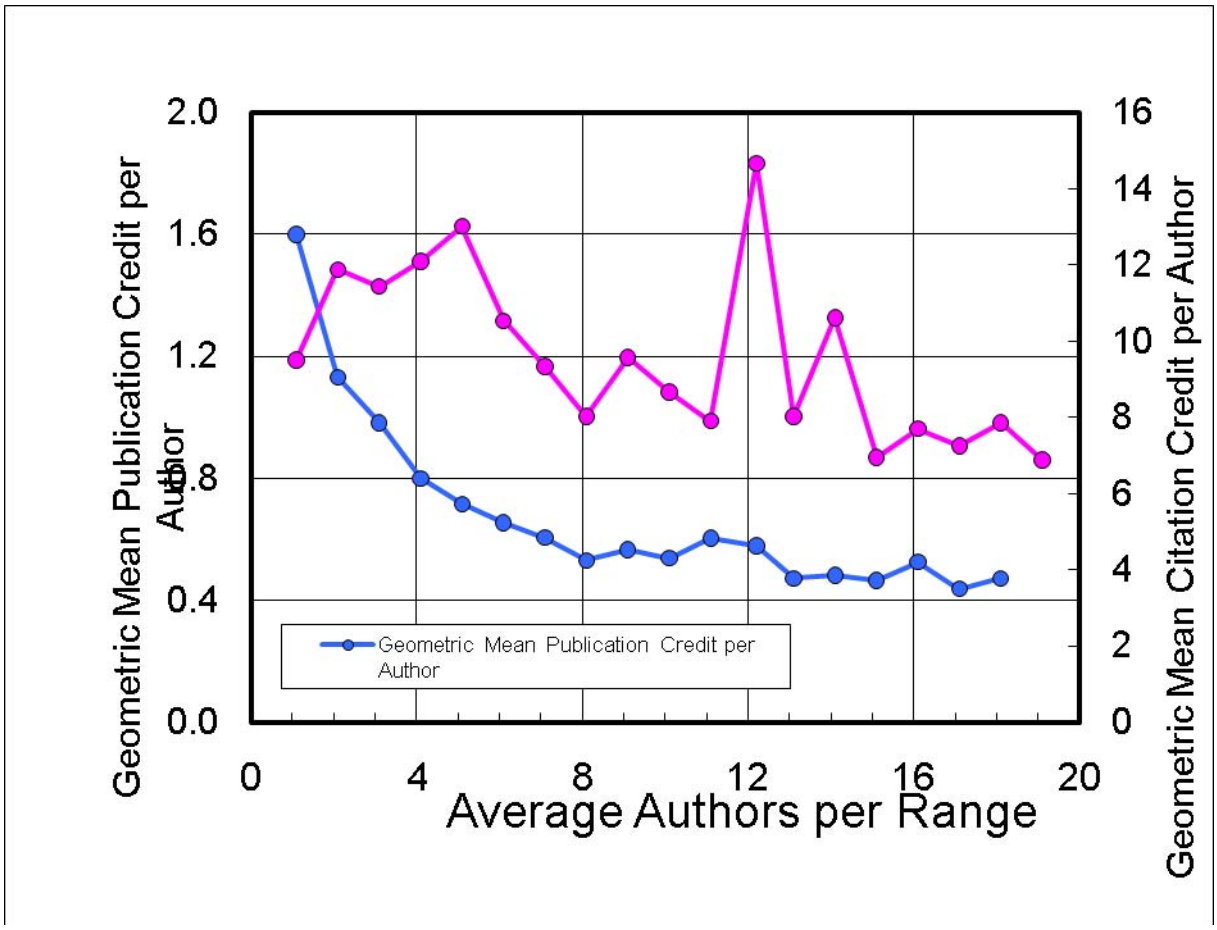

Figure 1: Geometric mean of productivity and citation obtained from the whole and fractional counting systems (small number of authors).

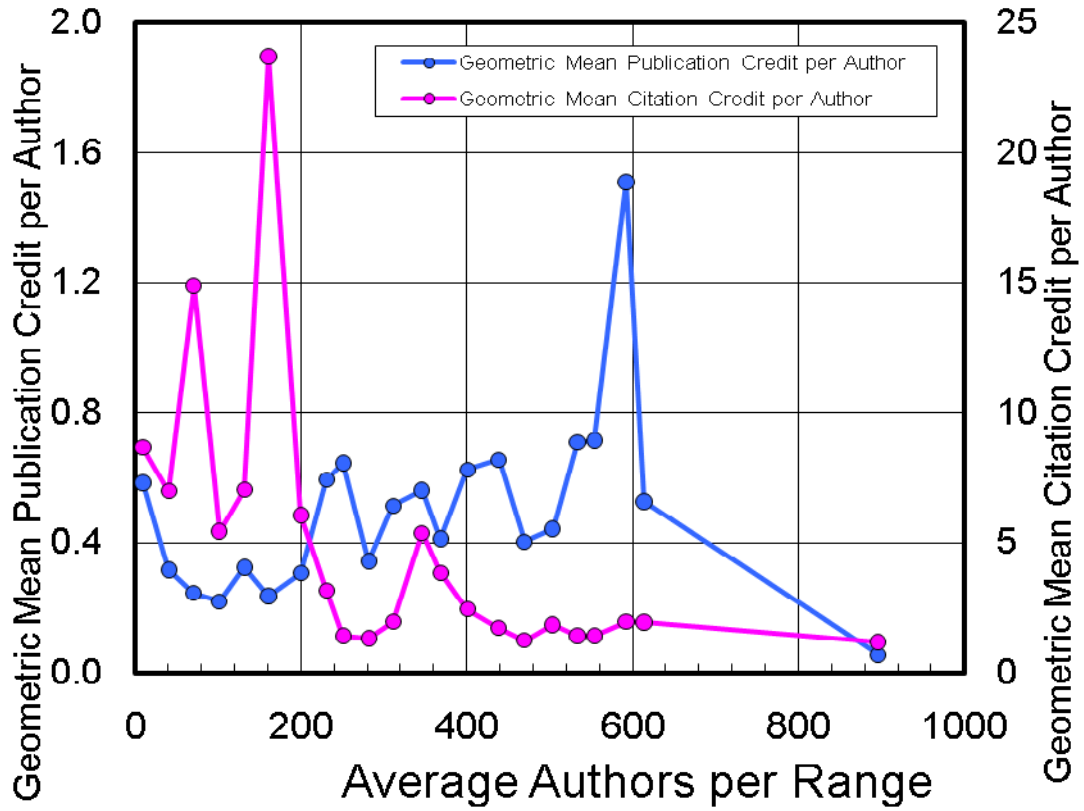

Figure 2: Geometric mean of productivity and citation obtained from the whole and fractional counting systems (large number of authors).

Although it is uncomfortable to realise that research findings can depend on the counting system it is imperative that citation analysis acknowledges its anomalies. Acknowledging anomalies is an important step towards addressing anomalies and addressing them will ultimately strengthen citation analysis. As regards which counting system is preferable we tend to agree with de Solla Price (1981) who recommended using the fractional counting system "in the absence of evidence to the contrary." But is there any evidence to the contrary? A justification for the whole counting system is that in peer review generally no adjustment is made for the number of authors; a justification for the fractional counting system is that in citation analysis all articles are usually given equal weighting. Faced with two justifiable systems that lead to different findings, one could adopt a compromise counting system that is between the whole and fractional systems. One possibility is the square-root system, in which the whole credit awarded to each article is the square-root of the number of authors. If the credit is shared equally amongst the authors, then an author of a paper with four authors would receive half the credit of the solo author of a paper, the total credit awarded to the paper with four authors would be double that awarded to a sole author of a paper. Ultimately, however, it is up to researchers to make a qualitative judgement about how much credit should be allocated to articles by co-authors, and this may vary by discipline, especially for disciplines with non-standard attribution practices. The results presented here make a clear case for the importance of such a decision for physicists.

Finally, the 183 articles in Physics_AstroNuclear with over 200 authors pose considerable problems for the effective evaluation of physics research. As described in the Appendix, 472

researchers from 60 institutions and 11 countries are authors of the same 57 articles (Cluster 1). This high level of international co-authorship of this cluster is likely to distort findings on international co-authorship in physics. On the basis of these 57 articles, each author receives an extremely high whole citation count of 1,386 and a low fractional citation count of less than 3. Each of these 472 authors has an h-index of at least 21 for 2004 (as they receive this from their 57 articles); outside of these 472 authors, only one author of Physics_AstroNuclear articles received an h-index of at least 21 for 2004 articles. Furthermore, as the authors are presented in alphabetical order, the list of authors does not indicate their relative contributions and so there is no possibility to resolve this issue with a non-standard citation counting method. If these authors are to be assessed for funding for their research, and some of the UK academics will be, then this issue urgently needs to be resolved.

Acknowledgements

This research is supported by the Economic and Social Research Council [grant reference: RES-000-22-4415]. We wish to thank Gertrude Levitt for her very helpful comments on the drafts.

References

Burrell, Q., & Rousseau, R. (1995). Fractional counts for authorship attribution: A numerical study. *Journal of the American Society for Information Science*, 46(2), 97–102.

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.

Frederiksen, L.F. (2004). Disciplinary determinants of bibliometric impact in Danish industrial research: Collaboration and visibility. *Scientometrics* 61(2), 253-270.

Gauffriau, M. & Larsen P.O. (2005). Counting methods are decisive for rankings based on publication and citation studies. *Scientometrics* 64(1), 85–93.

Glänzel W. & De Lange C. (2002). A distributional approach to multinationality measures of international scientific collaboration. *Scientometrics* 54(1), 75-89.

Glänzel, W. & Schubert A. (2001). Double effort = Double impact? A critical view at international co-authorship in chemistry. *Scientometrics* 50(2), 199-214.

Lee, S. & Bozeman, B. (2005). The Impact of Research Collaboration on Scientific Productivity. *Social Studies of Science* 35 (5), 673-702.

Hagen, N.T. (2010). Harmonic publication and citation counting: sharing authorship credit equitably – not equally, geometrically or arithmetically. *Scientometrics* 84(3), 785-793.

Levitt, J.M. & Thelwall, M. (2009). Citation levels and collaboration within Library and Information Science. *Journal of the American Society for Information Science and Technology* 60(3), 434-442.

Levitt, J.M. & Thelwall, M. (2010). Does the higher citation of collaborative research differ from region to region? A case study of Economics. *Scientometrics* 85(1), 171-183.

Lindsey D. (1980). Production and Citation Measures in the Sociology of Science: The Problem of Multiple Authorship. *Social Studies of Science* 10(2), 145-162.

Lotka, A. J. (1926). The Frequency Distribution of Scientific Productivity, *Journal of the Washington Academy of Science* 16, 317-23.

Pao, M.L. (1982). Collaboration in Computational Musicology, *Journal of the American Society for Information Science* 33(1), 38-43.

Pravdic, N. & Oluic-Vukovic, V. (1986). Dual Approach to Multiple Authorship in the Study of Collaborator and Scientific Output Relationship. *Scientometrics* 10(5/6), 259-80.

Price, D.J. de Solla & Beaver, D. (1966). Collaboration in an Invisible College, *American Psychologist* 21(11), 1011-18.

Price, D.J. de Solla (1981). Multiple Authorship. *Science* 212(4498), 986.

Sooryamoorthy R. (2009). Collaboration and publication: How collaborative are scientists in South Africa? *Scientometrics* 80(2), 419-439.

Vogel, E.E. (1997). Impact factor and international collaboration in Chilean physics: 1987-1994. *Scientometrics* 38(2), 253-263.
Wuchty, S., Jones, B.F. and Uzzi, B. (2007). The Increasing Dominance of Teams in Production of Knowledge. *Science* 316(5827), 1036-1039.
Zuckerman, H. (1967). Nobel Laureates in Science: Patterns of Productivity, Collaboration, and Authorship. *American Sociological Review* 32(3), 391-403.

Appendix

Figure 3 depicts the distribution of the articles with over 50 authors (excluding a single outlier that has 939 authors).

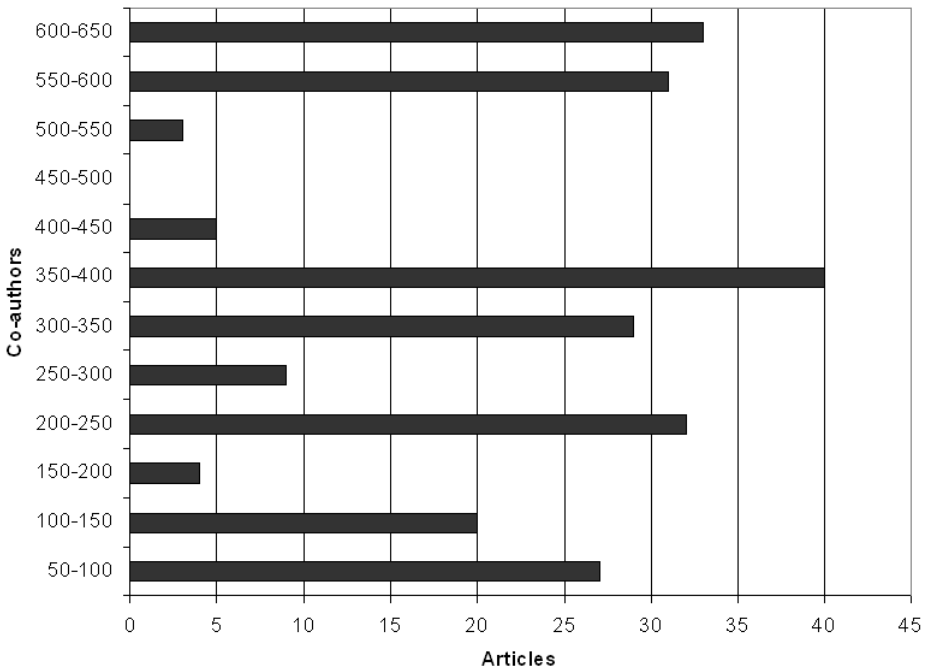

Figure 3. Number of articles with 50-650 co-authors.

The 183 articles with over 200 authors were divided into 14 clusters of authors and the data on each cluster presented in Table 3. The clusters were selected in such a way that each has a minimum of 100 researchers who authored every article in the cluster. The 183 articles usually present their authors in alphabetical order. The 472 authors who published every article in the largest cluster were affiliated to 60 institutions in 11 countries. The 57 articles were published in exactly two journals (Physical Review Letters and Physical Review D), articles include 'Direct CP violating asymmetry in B-0 -> K+pi(-) decays' and 'Study of the e(+)e(-)->pi(+)pi(-)pi(0) process using initial state radiation with BABAR'.

Table 3: Some bibliometric data on the 14 clusters in Physics_AstroNuclear.

Cluster number	*Articles in the cluster*	*Authors of every article in cluster*	*H-index of the cluster*	*Citations of the cluster*
1	57	472	21	1,386
2	20	192	12	314
3	19	291	11	240
4	16	246	11	628
5	15	215	12	1,149
6	15	235	10	499
7	11	237	11	302
8	9	139	8	516
9	7 *	295	6	302
10	5	272	4	85
11	5	372	5	469
12	3	189	3	185
13	1	349	1	282
14	1	237	1	86

*This cluster overlaps strongly with one of the clusters of 1 article.