



The Problem of Junior-Authored Papers in Constructing Citation Counts

J. Scott Long; Robert McGinnis; Paul D. Allison

Social Studies of Science, Vol. 10, No. 2 (May, 1980), 127-143.

Stable URL:

<http://links.jstor.org/sici?sici=0306-3127%28198005%2910%3A2%3C127%3ATPOJPI%3E2.0.CO%3B2-F>

Social Studies of Science is currently published by Sage Publications, Ltd..

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/about/terms.html>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at <http://www.jstor.org/journals/sageltd.html>.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is an independent not-for-profit organization dedicated to creating and preserving a digital archive of scholarly journals. For more information regarding JSTOR, please contact jstor-info@jstor.org.

● ABSTRACT

Citation counts are central to many studies in the sociology of science. Due to cost considerations and the format of the Science Citation Index, researchers often collect citations only to a scientist's first-authored publications (so called 'straight counts'), rather than to complete counts based on all published papers.

This paper examines the consequences of using straight counts to study the careers of a sample of PhD biochemists. It is found that even though these two measures are highly correlated, there is a significant chance of making substantive errors when straight counts are substituted for complete counts. The problem is found to be more severe when citations are the dependent variable rather than an independent variable. It is argued that our findings can be generalized to other fields where multiple authorship is common. Accordingly, extreme care must be used in interpreting results based on straight citation counts.

The Problem of Junior-Authored Papers in Constructing Citation Counts

J. Scott Long,
Robert McGinnis and
Paul D. Allison

Citation counts have become standard measures of productivity in studies of the scientific career. Ideally a citation measure would include citations to all papers a scientist authors, whether he or she is the first author, the last author, or someplace in between. Unfortunately, *Science Citation Index (SCI)* — the standard source of citation counts — does not list citations to individuals, but rather to articles. Each paper is listed only under the name of the first author of the paper. Thus, if a scientist writes numerous papers, but is never the first author, his name will not be found in *SCI*. Citations to a scientist's junior-authored papers can be found only if the name of the senior author is known. While this information usually

can be found, recording it substantially increases the cost of citation coding.

In most prior research the choice has been to forego the collection of citations to all published papers (hereafter referred to as *complete counts*) in favour of *straight counts*, that is, counts based only on first-authored papers. For example, Allison and Stewart, Bayer and Folger, Cole and Cole, Hagstrom, Hargens and Hagstrom, and Reskin have used only citations to first-authored articles.¹ Gaston and Long are among studies that have used citations to an author's entire work.²

While considerations of cost have led many authors to use straight counts rather than complete counts, substantial concern has been expressed over the possibility that bias may be introduced by the use of straight counts. Reskin has suggested that straight counts may systematically underestimate the number of citations that young scientists receive.³ Porter thought that straight counts would benefit theorists, while complete counts would better represent the productivity of experimentalists.⁴ He also noted that both the incidence of co-authorship and the number of authors per paper has increased and varies with the 'hardness' of a field, thus possibly introducing bias into longitudinal and cross-disciplinary studies. Garfield warns:

How greatly this inaccuracy [using straight counts to estimate complete counts] distorts relative citation measurements is a matter of considerable debate. . . . As long as this uncertainty persists, the only fair way of developing relative citation counts is to compile the performance of all the published material that is listed on a comprehensive bibliography.⁵

This statement is in strong contrast to one by Cole and Cole. In their early work involving citation analysis they claim that it is possible to use straight counts with reasonable confidence.⁶

The purpose of this paper is to assess the relative merits of straight and complete counts. First, how much error is introduced into substantive analyses by the use of straight counts rather than the presumably more accurate complete counts? Second, how much more costly are complete counts? We begin by describing the data base that will be used.

The Data Base

The analysis presented below is based on the population of male biochemists who obtained their doctorates in fiscal years 1957, 1958, 1962 and 1963. Female doctorates were excluded due to their relatively small number and the difficulty of obtaining complete information on those who did obtain their degrees. Of the population obtaining degrees in this period ($n = 668$), complete information was obtained for 83 percent ($n = 557$). Investigations on the 17 percent for whom information was not complete suggested that most of these PhDs either dropped out of biochemistry completely or were foreign born and returned to their native countries to pursue their careers.

Biographic and work history information was coded from the 10th, 11th and 12th editions of *American Men and Women of Science (AMWS)*.⁷ This was supplemented with information concerning job characteristics from the Carter, Roose and Andersen, and Astin reports.⁸ Information about the scientists' mentors (dissertation supervisors) was obtained from university records, *Dissertation Abstracts*, *AMWS*, and the *Directory of Graduate Research*.⁹

The Collection of Bibliographic Information

Bibliographic information for our cohort members was collected using *Chemical Abstracts (CA)* and *Science Citation Index (SCI)*. *CA* was used to obtain a complete list of a scientist's publications for the years 1955 to 1972; *SCI* was then used to count citations to these articles. This procedure resolved two basic problems: first, it was possible to obtain a list of a scientist's publications which included junior-authored papers, thus allowing the collection of complete citation counts; second, the homonym problem could be dealt with more effectively.

The homonym problem is the difficulty of distinguishing a scientist's articles from those of other scientists with similar names. This is a particularly difficult problem in *SCI*, where only the initials of a scientist are listed. By using *CA* to obtain a list of publications, the homonym problem is transferred from *SCI* to *CA*, where it can be more readily resolved. First, when using *CA* the homonym problem involves distinguishing only among chemists with similar

names, whereas in *SCI* one must distinguish among scientists in all fields. So, even if scientists were to be identified only on the basis of their last name and initials, the homonym problem would be significantly reduced by using *CA*. But the problem is further reduced by using *CA*, since an author's full name (that is, first name, middle name and last name) is generally provided. For a majority of cases this immediately solves the homonym problem. In cases where only the first name or initials are listed in *CA*, the institutional affiliations of the authors, which are listed in *CA*, can be compared to work histories coded from *American Men and Women in Science*.

For 85 percent of the more than 13,000 articles coded from *CA*, a complete match of the last name, first name and middle name (or initial) was possible. For an additional eight percent, the last name and both initials were matched. The remaining seven percent, which had only a single initial that matched, were excluded from our set of articles. It should be noted that nearly all of the excluded articles belonged to cohort members for whom we could not find biographic information in *AMWS* to aid in the verification process. Since these scientists were excluded from the analysis anyway, our set of articles should be nearly complete and quite accurate.

Using the 1961, 1964, 1966, 1968, 1970, 1972, and 1974 volumes of *SCI*, both the total number of citations to each article and the number of self-citations were coded for each article obtained from *CA*. Self-citations to an article were defined as citations made by either our cohort member or by the first author of the given article. Citations by authors other than the first and/or cohort author were missed in the counts of self-citations due to practical limitations in the coding operation. For purposes of analysis all known self-citations were excluded. Since coverage by *SCI* and *CA* increased during the period covered by our study, counts were standardized within years of the PhD. After standardization the scale of the measures was adjusted back to the scale of the raw counts and the minimum value was reestablished as zero.¹⁰

Substantive Consequences of Straight versus Complete Counts

Researchers who substitute straight counts for complete counts are implicitly assuming that while straight counts necessarily underestimate complete counts, the differences between straight

counts and complete counts are unrelated to other variables which either affect or reflect the scientific career. To the extent that this assumption is inaccurate, analyses which employ straight counts may obtain misleading results. The following analyses attempt to assess this possibility. First the effect of age on author position is examined. Second, the different results obtained when straight counts are substituted for complete counts in regression analyses are shown. Third, a proposal for estimating complete counts from straight counts is considered. And finally, the financial costs of collecting complete counts are discussed.

A possible source of bias is professional age.¹¹ If we examine consecutive three-year periods beginning the year before the granting of the doctorate, we find that in the first period 39 percent of the publishing scientists had at least half of their publications excluded from consideration when straight citation counts were made. For the three-year period ending four years after the PhD, 42 percent had at least half of their publications excluded; for the period ending seven years after the degree, 48 percent; and for the period ending ten years after the degree, 61 percent. Similarly, in the first three-year period, 16 percent of the publishing scientists had all of their publications excluded when straight counts were made; in the period ending in year four, 16 percent again; in the period ending in year seven, 17 percent; and in the final period, 21 percent. This trend of increasing possibility of bias in straight counts is also suggested by Table 1, which presents frequencies of publications and citations over time.

Research by Bayer and Folger, Reskin, and Porter¹² suggests that scientists are junior authors less frequently as they grow older. Our data present evidence to the contrary: junior-authorship increases with professional age. Accordingly, the problem of using straight counts may be greater later in a scientist's career, not earlier.

Table 1 shows that the average number of papers in a three-year period increases with professional age, from 1.9 in the period ending the year after the degree (year one) to 4.3 in the period ending ten years after the degree. The average number of first-authored papers, however, increases far more gradually, from 1.2 in year one to 2.0 in year seven, declining to 1.9 in year ten. Further, the proportion of first-authored papers as percent of total papers steadily decreases from 64 to 44 percent. Similar figures for total or cumulative publication counts are presented in the lower portion of the table; these figures are necessarily dampened due to their

TABLE 1
All of a Scientist's Publications Compared to His First-Authored Publications by Years Since PhD

	Years Since Doctorate			
	Year 1	Year 4	Year 7	Year 10
Mean Total Publications in 3-year Period Ending in Given Year	1.93 (1077)	3.17 (1763)	3.66 (2039)	4.33 (2412)
Mean First-Authored Publications in 3-year Period Ending in Given Year	1.25 (694)	1.89 (1052)	2.00 (1112)	1.93 (1072)
First-Authored Publications as Percent of Total Publications	64.4	59.7	54.6	44.4
Mean Total, Cumulative Publications Through Given Year	1.93 (1077)	5.10 (2840)	8.76 (4879)	13.09 (7291)
Mean First-Authored Cumulative Publications Through Given Year	1.25 (694)	3.14 (1746)	5.13 (2858)	7.06 (3930)
First-Authored Publications as Percent of Total Publications	64.4	61.5	58.6	53.9

Note: Results are based on population of 557 male biochemists who obtained their degrees between 1957 and 1963. Figures in parenthesis indicate the total number of publications represented by the mean figure.

cumulative nature. What these results show is that for year one straight citation counts would be based on 64 percent of the total number of articles written by the average scientist, and by year ten on 54 percent of the total number of publications. But since articles in biochemistry have a rather short citation half-life, the effective figure is probably closer to that for the three-year totals, which drops to 44 percent in year ten.

Comparable results were found for citation counts. The average straight citation count for a scientist in the three-year period ending the year of the PhD is 51 percent less than the average complete count during that period. By the period ending in the fourth year after the degree, this percentage increases to 62, at which point it steadily declines to 56 percent in the period ending in year seven, and to 42 percent in the period ending in year ten. In short, from the fourth year on, straight counts represent an increasingly small proportion of a scientist's complete citations. Contrary to what might be expected, there is a greater possibility of introducing error by using straight counts later in the career than earlier.

How do these factors affect the citation counts themselves? In Table 2 the correlations between complete counts and straight counts are presented for citation counts based on work in the previous three-year period and based on all past work. For the three-year counts the correlations are fairly constant at about 0.8. For total counts, the correlations are somewhat larger with values around 0.85. These correlations would generally be considered to be quite high, but still the key question is whether the 25 to 36 percent of unexplained variation is random measurement error or systematic bias.

Table 3 presents a regression analysis of the effects of various characteristics of the doctorate and the productivity of the scientist on the prestige of the first academic job.¹³ In comparing the results obtained using complete counts with those obtained using straight counts (compare computation 1 with 2, and 3 with 4), we note that the regression coefficients for the productivity measures are attenuated and that the significance levels decrease when using straight counts. The other coefficients in the equations using straight counts increase slightly. While the differences between the equations using complete counts and straight counts are not great, in the case of total counts the effect could make a difference in the substantive conclusions drawn.

TABLE 2
Correlations Between Complete Citation Counts, Straight Citation Counts, and Estimates of Complete Citation Counts

Correlations Between Complete and Straight Citation Counts				
	Year 1	Year 4	Year 7	Year 10
Counts Based on All Past Articles	0.796	0.848	0.876	0.863
Counts Based on Articles in the Past Three Years	0.796	0.850	0.804	0.754
Counts Based on Articles in the Past Three Years for Sample of 134	0.737 ^a		0.866 ^b	
Correlations Between Complete and Estimates of Complete Citation Counts				
Counts Based on All Past Articles	0.853	0.873	0.885	0.901
Counts Based on Articles in the Past Three Years	0.853	0.841	0.828	0.796
Counts Based on Articles in the Past Three Years for Sample of 134	0.788 ^a		0.783 ^b	

Note: Except as noted, results are based on population of 557 male biochemists who obtained their degrees between 1957 and 1963. Estimates of total citation counts were computed as follows:

$$\text{ESTIMATE} = \frac{(\text{Total papers}) \times (\text{Straight citation count})}{(\text{First-authored papers})}$$

- a. Based on 3-year period ending one year after start of first academic, tenure-track job.
- b. Based on 3-year period ending six years after start of first academic, tenure-track job.

TABLE 3
Results Obtained by Using Complete, Straight and
Estimated Complete Counts in Regressions Estimating
the Prestige of the Scientist's First Academic Position

Productivity Measure Used as Independent Variable		PHD	MENT	BA	CIT	R²
1. Complete	b	0.362	0.192	0.146	0.126	0.307
Total Citations	t	4.61	2.38	1.93	1.65	
For Year Job Was Obtained	r	0.464	0.372	0.249	0.247	
2. Straight	b	0.363	0.201	0.155	0.075	0.297
Total Citations	t	4.57	2.47	2.04	0.97	
For Year Job Was Obtained	r				0.203	
3. Complete	b	0.364	0.197	0.146	0.114	0.304
3-Year Citations	t	4.63	2.46	1.93	1.50	
For Year Job Was Obtained	r				0.255	
4. Straight	b	0.366	0.201	0.145	0.084	0.298
3-Year Citations	t	4.62	2.49	1.88	1.08	
For Year Job Was Obtained	r				0.204	
5. Estimated	b	0.362	0.207	0.157	0.063	0.296
Complete 3-Year	t	4.55	2.56	2.06	0.82	
Citations for Year Job Was Obtained	r				0.186	

Note: Regressions are based on the sample of 134 male biochemists employed in rated graduate departments. Dependent variable is Roose-Andersen prestige score of the institutional location. Item identifications are:

PHD = origin prestige (Cartter prestige of PhD department if nonfellow or if nonacademic fellow, and Roose-Andersen bioscience prestige score if fellow);

MENT = square root of five-year citations to mentor;

BA = selectivity of baccalaureate institution;

CIT = measure of citations (as indicated in left hand column of table);

b = the standardized regression coefficient;

t = the t-value for the test of the regression coefficient; and

r = the simple correlation between the independent variable and the dependent variable.

In Tables 4 and 5, larger differences are found. In both tables alternative measures of productivity are treated as dependent variables with characteristics of the scientist's education and (for Table 5) the prestige of his academic location as the independent variable.

For citation counts made the year after the doctorate (Table 4), we see that there are several differences between the results obtained with straight counts and with complete counts. The major difference is that the effects of the mentor's prestige (as indicated by the number of citations he or she received in a five-year period near the student's doctorate) are greatly attenuated. The effects of the prestige of the doctorate and of collaborating with the mentor increase slightly in the regression for straight counts, and the effect of the selectivity of the baccalaureate is substantially increased. These findings reflect the results of regressing (regression not shown) the percentage error between the straight and complete counts, with complete counts being considered the base, on characteristics of the doctorate, the mentor and the level of productivity. In this regression the largest coefficient was that from the prestige of the mentor ($b = 0.233$, $p = 0.01$), with the coefficients from the prestige of the doctorate ($b = 0.162$, $p = 0.01$) and the selectivity of the baccalaureate ($b = 0.090$, $p = 0.10$) being the next largest. In other words, the attenuation of the effect of mentor's prestige is a consequence of the fact that straight counts more seriously underestimate complete counts when the individual has a prestigious mentor.

When we examine the regressions for productivity in the sixth year after the start of the first academic job (Table 5), somewhat different results are obtained. The effect of the current affiliation is attenuated in the regressions for straight counts, while the effects of the mentor's prestige are inflated. The other coefficients are too small to be of concern.

It seems clear that misleading interpretations regarding the influence of the mentor would be made if straight counts were used. At the time of the doctorate the effect of the prestige of the mentor would be underestimated, while later in the career it would be overestimated. One would incorrectly conclude that the effect of the mentor is smaller but more persistent than it really is.

One solution to the problem of straight counts is to estimate complete citation counts from the straight counts using information on the proportion of times a scientist is the first author. This

TABLE 4
Regressions on Complete and Straight Citation Counts Measured
the Year After the Receipt of the Doctorate

Productivity Measure Used as Dependent Variable		PHD	MENT	COLLAB	BA	R ²
1. Complete Citations	b	0.062	0.255	0.286	0.128	0.244
	t	0.97	3.94	4.83	2.19	
	r	0.203	0.357	0.345	0.175	
2. Straight Citations	b	0.093	0.172	0.301	0.162	0.206
	t	1.45	2.62	5.04	2.73	
	r	0.206	0.294	0.345	0.202	

Note: Results are based on 239 male biochemists whose first job was in a rated graduate institution. Item identifications are as in Table 3, with the addition of:

COLLAB = collaboration with mentor (defined as one if any article published by the year after doctorate was coauthored with the mentor, zero otherwise).

method was proposed by Roy.¹⁴ The formula for estimating the complete citation count at a given time *t* is:

$$\text{ESTIMATE} = \frac{(\text{Total papers through time } t) \times (\text{Straight citation count at time } t)}{(\text{First-authored papers through time } t)}$$

This formula was used to compute estimated complete citation counts for the scientists in our sample. In Table 2 we note that the correlations between the complete counts and the estimates increase from 0.85 in year one to 0.90 in year ten, when citations to all past work are considered. For three-year counts the correlations decrease from 0.85 to 0.79. Only in year ten for total counts is the correlation smaller between estimates and complete counts than between complete counts and straight counts. This suggests that Roy's method may be a viable approach for obtaining improved productivity measures at relatively modest cost. We note, however, that for the sample of 134 the correlation between the complete and estimated counts is only 0.783, compared to a correlation of 0.866 between complete and straight counts.

TABLE 5
Regressions on Complete, Straight and Estimated Complete
Citation Counts Measure Six Years After the Receipt of
1st Academic Job

Productivity Measure Used as Dependent Variable		PRST	PHD	MENT	BA	R²
1. Complete Total Citations	b	0.312	0.010	0.186	0.051	0.194
	t	3.31	0.11	2.12	0.62	
	r	0.399	0.227	0.314	0.160	
2. Straight Total Citations	b	0.255	0.032	0.247	0.092	0.215
	t	2.74	0.35	2.86	1.13	
	r	0.384	0.249	0.368	0.199	
3. Complete 3-Year Citations	b	0.344	0.095	0.091	0.038	0.205
	t	3.67	1.04	1.05	0.47	
	r	0.431	0.292	0.259	0.151	
4. Straight 3-Year Citations	b	0.272	0.073	0.140	0.052	0.166
	t	2.84	0.79	1.57	0.63	
	r	0.371	0.256	0.275	0.152	
5. Estimated Complete 3-Year Citations	b	0.187	0.155	0.147	0.086	0.167
	t	1.96	1.67	1.65	1.03	
	r	0.336	0.305	0.285	0.177	

Note: Results are based on the sample of 134 male biochemists employed in rated departments. Item identifications are as in Table 3, with the addition of:

PRST = prestige of institutional location (Roose-Andersen bioscience prestige score).

Accordingly for this sample of academic scientists the estimate may not be as useful as it first appears.

Indeed, in Table 3 we find that the coefficient for the estimated complete counts is smaller than that for the straight counts, although the coefficients are small in all cases. In Table 5 the estimated counts are also shown to introduce bias into the results, albeit a different bias than obtained when using straight counts. In this case (see computation 5) the effect of the institutional location is attenuated while the effect of the doctoral prestige is inflated; the effect of the mentor's citation level, however, remains about the same.

Conclusions

Overall we find that while scientists increase their average number of publications as they age, the proportion of papers in which they are the first author decreases steadily with time. Thus, the problem of using straight counts is *potentially* greater for professionally older scientists. The key question in the evaluation of straight counts is whether their use will bias the results of analyses in which they are used. Our findings suggest that when straight counts are employed as *independent variables* they attenuate the coefficients only slightly in the regression. In equations where productivity is a more significant factor, however, this attenuation may be more severe. When productivity is the *dependent variable*, the substitution of productivity measures based on straight counts for those based on complete counts can bias the findings in substantively important ways. In our example the effects of institutional location were attenuated. The effects of the mentor's eminence were inflated later in the career, while being attenuated at the time of the doctorate.

The alternatives to using straight counts are to either estimate complete counts (as Roy suggested), or actually to collect complete counts. Applying Roy's method for estimating complete counts, we found that the correlations between the estimates and the complete counts were generally, but not always, higher than the correlations between complete counts and straight counts. Still, in substantive examples the results obtained using the estimates were somewhat inferior to those obtained with straight counts.

If complete counts are collected, the cost of data collection is significantly increased. As described above, it is necessary first to consult an abstracting service and code information on each article. It is

not sufficient to code only the name of the first author since additional information is necessary to match the article with information in *SCI*. While various methods of collecting complete counts are possible, our cost estimates may be useful to others.

On the basis of our study of biochemists, we estimate that to count citations to all articles rather than just those to articles in which the cohort member was the first author increased the coding time required by a factor of roughly 2.6. While this is a substantial increase in cost, it is important to realize exactly what this increase includes. First, by including articles by junior authors, the number of articles for which citations were coded nearly doubled. When this is taken into account, the per-article cost was increased only by a factor of 1.4.

This added per-article cost had several benefits, besides aiding in the collection of complete citation counts. Of particular importance was that the author's institutional affiliation could be coded, which helped to deal with the homonym problem, as noted above. Second, the subject area of the article, as indicated by the abstracting service, could be coded during this procedure. Additional information, such as the number of coauthors on a paper, could also be collected with only a trivial increase in cost.

Still, the increase in cost is substantial and the researcher must decide whether the benefits merit the cost. The substantive error introduced by substituting straight counts for complete counts is critical in making this decision. Even though the two measures of productivity are highly correlated, our study has shown that when productivity is the dependent variable there is a significant chance of making a substantive error when straight counts are substituted for complete counts.

While our results are limited to biochemists, research presented by Meadows suggests implications for other fields.¹⁵ He notes that 42 percent of the mathematicians and statisticians studied, 83 percent of the experimental biologists and 93 percent of the physicists reported working with graduate students on research; 21 percent of the experimental biologists, 25 percent of the mathematicians and 50 percent of the physicists reported working with faculty colleagues.¹⁶ These differences are also reflected in figures on the percentage of multi-authored papers in various fields: in chemistry 83 percent are multi-authored; in biology 70 percent; in physics 67 percent; in mathematics 15 percent; and in history 4 percent.¹⁷ These data suggest that problems in substituting straight counts for complete counts will not be limited to biochemistry. In physics, biology, and chemistry there is every reason to believe that similar

problems will occur. In mathematics and history the likelihood of bias seems remote. Still, further research is necessary before our results can be safely generalized. Until more evidence is available, we would recommend following Garfield's advice: '[T]he only fair way of developing relative citation counts is to compile the performance of all the published material that is listed on a comprehensive bibliography.'¹⁸

● NOTES

This research was made possible by support from *NIH* contract N01-0D-4-2175.

1. P. D. Allison and J. A. Stewart, 'Productivity Differences Among Scientists: Evidence for Accumulative Advantage', *American Sociological Review*, Vol. 39 (1974), 596-606; A. E. Bayer and J. Folger, 'Some Correlates of a Citation Measure of Productivity in Science', *Sociology of Education*, Vol. 39 (1966), 381-89; J. R. Cole and S. Cole, *Social Stratification in Science* (Chicago: The University of Chicago Press, 1973); W. O. Hagstrom, 'Departmental Prestige and Scientific Productivity', paper presented at the 63rd Annual Meeting of the American Sociological Association, Boston, Mass., 27 August 1968; L. L. Hargens and W. O. Hagstrom, 'Sponsored and Contest Mobility of American Academic Scientists', *Sociology of Education*, Vol. 40 (1967), 24-38; B. F. Reskin, 'Scientific Productivity and the Reward Structure of Science', *American Sociological Review*, Vol. 42 (1977), 491-504; Reskin, 'Academic Sponsorship and Scientists' Careers' (unpublished paper, Indiana University, 1978).

2. J. Gaston, *The Reward System in British and American Science* (New York: Wiley, 1978); J. S. Long, 'Productivity and Academic Position in the Scientific Career', *American Sociological Review*, Vol. 43 (1978), 889-908.

3. Reskin (1977), op. cit. note 1, 495.

4. A. L. Porter, 'Citation, Analysis: Queries and Caveats', *Social Studies of Science*, Vol. 7 (1977), 257-67.

5. E. Garfield, *Citation Indexing — Its Theory and Application in Science, Technology, and Humanities* (New York: John Wiley & Sons, 1979), 242-43.

6. Cole and Cole, op. cit. note 1, 35.

7. Cattel Press, *American Men and Women of Science* (New York: R. R. Bowker Co.).

8. A. M. Cartter, *An Assessment of Quality in Graduate Education* (Washington, DC: American Council on Education, 1966); K. D. Roose and C. J. Andersen, *A Rating of Graduate Programs* (Washington, DC: American Council on Education, 1970); A. W. Astin, *Predicting Academic Performance in College* (New York: The Free Press, 1971).

9. Xerox Corporation, *Dissertation Abstracts* (Ann Arbor, Mich.: University Microfilms); Cattel Press, op. cit. note 7; American Chemical Society, *Directory of Graduate Research* (Washington, DC: American Chemical Society).

10. See Long, op. cit. note 2, Appendix, for details on the standardization.

11. One possible source of difficulty in using straight counts is that co-authors may arrange their names alphabetically. To the extent that this is true, some scientists will consistently be first authors while others will rarely have that position, simply because of the first letters of their last names. Research by Cole and Cole (op. cit. note 1, 28) and Porter (op. cit. note 4) suggests that this is not a problem. Studies by D. Lindsey ('Problems of Measurement in the Sociology of Science: Taking Account of Collaboration', paper presented at the Annual Meeting of the American Sociological Association, Chicago, September 1977, since published as 'Production and Citation Measures in the Sociology of Science: The Problem of Multiple Authorship', in this issue, *Social Studies of Science*, Vol. 10 [1980], 145-62), and E. Rudd ('The Effects of Alphabetic Order of Author Listing on the Careers of Scientists', *Social Studies of Science*, Vol. 7 [1977], 268-69), however, suggest that one's standing in the alphabet may affect one's likelihood of being the first author on collaborative papers. While these results are not consistent, in none of the studies was the alphabetic effect large.

12. Bayer and Folger, op. cit. note 1; Reskin (1977), op. cit. note 1; Porter, op. cit. note 4.

13. See Long, op. cit. note 2, for further details.

14. R. Roy, 'Approximating Total Citation Counts from First-Author Counts and Total Papers', working paper, cited by Garfield, op. cit. note 5, 242.

15. A. J. Meadows, *Communication in Science* (London: Butterworths, 1974).

16. Ibid., 196.

17. Ibid., 197.

18. E. Garfield, op. cit. note 5, 243.

J. Scott Long is Assistant Professor in the Department of Sociology at Washington State University. Long's current research presently proceeds on two paths: 1) the process of stratification in the scientific career; and 2) multivariate statistical techniques.

Paul D. Allison is Assistant Professor in the Department of Sociology at Cornell University. With J. Scott Long he is investigating the relationship between university departmental prestige and the productivity of individual scientists. He also is developing new methods for the analysis of event history data.

Robert McGinnis is Director of the Research Programme on Social Analyses of Science Systems at Cornell. He is doing research on networks of basic and applied research, researchers, and research organizations in agricultural science.

Author's addresses: (JSL) Department of Sociology,
Washington State University, Pullman, Washington 99164, USA;
(RMcG and PDA) SASS, Cornell University, 323 Uris Hall,
Ithaca, New York 14853, USA.