Appendix F. Analysis of the Relative Citation Index

We sought to examine whether the number of citations per publication at an institution (i.e., the relative citation index, or RCI) was related to various characteristics of the institution. The relative citation index is a measure of the quality or influence of publications, although there are reasons other than these that can result in more or fewer citations per publication. We defined the relative citation index for a given year as the ratio of number of citations that occur two years in the future to the count of publications occurring in the three year period ending in the given year. For example, we defined the relative citation index for 1992 to be the number of citations in 1994 measured using whole counts in an expanding journal set to all publications in 1990, 1991, and 1992 measured using whole counts.

The range of the relative citation index over institution-years is 0.59 to 19.9; 95% of the values at the institution-year level are between 1.7 and 9.9; the mean value is 4.89; the median is 4.47; and the standard deviation is 2.27. A very large proportion (94.3%) of the variance in the relative citation index is attributable to differences between institutions. The relative citation index varies by field. The average number of citations per publication is largest in the biology-life-agricultural sciences (6.71) and medical sciences (5.11); the average number of citations per publication is intermediate in engineering-math-physical sciences (3.66) and social sciences and psychology (2.37) and is relatively small in computer science (1.17). The correlation of the relative citation index and whole count publications is 0.58; the correlation with whole count citations is 0.69. This suggests that some of the same factors that account for increasing publications and citations may be associated with higher levels of citations per publication.

In regression analyses that we conducted, the dependent variable was the relative citation index. The data set was aggregated to the institution-year level. Financial variables were deflated using the GDP deflator; personnel variables were lagged by one year and financial variables by two years. Analyses were conducted using stepwise linear regression. We typically retained any variable that, when entered into the regression, increased r-squared by at least 0.01. All such variables were highly statistically significant when entered.

The set of academic R&D expenditures variables in their order of entry into the regression and the cumulative r-squared values after entry were as follows: 1) federally financed academic R&D expenditures (0.328), 2) non-federally financed academic R&D expenditures (0.373), 3) other funding sources for academic R&D expenditures (0.469), 4) total basic research expenditures (a component of total academic R&D expenditures, 0.484), and 5) institution financed unreimbursed indirect costs (0.496). The remaining academic R&D expenditure and research equipment variables, total academic R&D expenditures and research equipment expenditures, federally financed basic research and research equipment expenditures, industry, state/local government, and institutionally financed academic R&D expenditures, and institutionally financed organized research expenditures, did not meet a 0.01 threshold. The coefficients for four of these variables were positive. The coefficients per $1M for these variables, in their order of entry, were 0.01, -0.04, 0.14, 0.016, and 0.055. We note that since academic R&D funded by other

sources is non-federal, the net coefficient for this type of funding is 0.10. Thus, research funded by foundations and the institution itself appear to increase the number of citations per publication by the most, and research funded by other non-federally funded sources (primarily commercial sources and state/local governments) decreases the number of citations per publication.

The set of enrollment variables in their order of entry into the regression and the cumulative r-squared values after entry were as follows: 1) fall enrollment of undergraduates (0.43), 2) and fall enrollment of graduate students (0.174). These fall enrollment variables were institution-wide rather than S&E specific; the number of S&E graduate students did not enter the model. The coefficient for graduate students was 0.36 per 1,000 students. The coefficient for undergraduates (-0.11 per 1000 students) was negative, possibly a reflection that institutions with more undergraduate students have lower production of publications which is associated with a smaller relative citation index.

The set of postdoctoral count variables in their order of entry into the regression and the cumulative r-squared values after entry were as follows: 1) total postdoctorates (0.366), 2) postdoctorates supported by federal traineeships (0.377), and 3) postdoctorates with M.D.s supported by federal traineeships (0.397). The remaining variables did not meet a 0.01 threshold. The coefficients per 1,000 postdoctorates for these variables were 3.4, -60.0, and 46.2. The net effect is therefore -56.6 for postdoctorates without M.D.s supported by federal traineeships, -10.4 for postdoctorates with M.D.s supported by federal traineeships, and 3.4 for all other postdoctorates. The negative effects may reflect the emphasis of federal traineeships on training or other activities rather than research, which lowers production of articles. Institutions with a lower production of articles are associated with a smaller relative citation index.

The set of non-faculty research staff count variables in their order of entry into the regression and the cumulative r-squared values after entry were as follows: 1) non-faculty research staff with M.D.s (0.253), and 2) non-faculty research staff without M.D.s (0.264). The coefficients per 100 non-faculty research staff were 14.7 and 0.6, respectively, indicating that non-faculty research staff with M.D.s result in higher citations per publication. This is may reflect the larger number of citations per publication in the medical and biological-life-agricultural sciences fields.

The set of faculty count variables in their order of entry into the regression and the cumulative r-squared values after entry were as follows: 1) associate professors (0.014), and 2) full professors (0.148). Instructors, lecturers, assistant professors and other ranks did not meet the 0.01 threshold. The coefficients per 100 faculty for the two ranks entering the regression were -0.91 and 0.59, respectively. We do not know why larger numbers of associate professors are associated with fewer citations per publication. Possible explanations could include associate professors being more concentrated in fields with fewer citations per publication, being funded more often by sources associated with fewer citation per publication (i.e., commercial or state/local governmental sources), having classes with larger numbers of undergraduates, being more intent on publishing to acquire promotion to full professorship even though each publication may not be as influential, being more highly concentrated in institutions with unmeasured

characteristics that influence publications, etc. Associate professors with more citations per publication may also be promoted more quickly to full professorship than associates with fewer citations per publication.

The set of degrees awarded variables in their order of entry into the regression and the cumulative r-squared values after entry were as follows: 1) S&E doctorates from the NSF Survey of Earned Doctorates (0.090), and 2) S&E BA/BS degrees (0.242). S&E doctorates from the IPEDS Completions Survey, and S&E master's degrees awarded did not meet the 0.01 incremental threshold. The coefficient per 1,000 degrees for BA/BS degrees was negative (-1.4) and the coefficient for Ph.D. degrees was positive (1.4), consistent with the finding for number of enrolled graduate students. The two doctoral degree variables had a correlation of 0.973.

The set of Carnegie classification variables in their order of entry into the regression and the cumulative r-squared values after entry were as follows: 1) R-1 (0.173), 2) Med (0.306), and 3) M-1 (0.325). The coefficient for MED and R-1 were positive (2.7 and 2.3, respectively) and the coefficient for M-1 was negative (-1.6). The negative coefficient may reflect the emphasis on relatively smaller doctoral programs in M-1 institutions and their emphasis on careers outside of academia. The large coefficient for medical schools may reflect the concentration of citations to the medical literature occurring from two to four years after publication.

A regression with the dependent variables being the type of control was also conducted. Public control entered the regression with an r-squared of 0.174 and a coefficient of -2.1.

We performed a stepwise regression utilizing all variables originally entered into any of the above linear regressions. In addition, we noticed a slight time trend in the residuals, so we added a time variable (coded 1 to 14 for years 1988 through 2001). The variables that increased r-squared by 0.01 or greater are listed in Table F-1 below, along with the cumulative r-squared, the regression coefficient, and the standard error of estimation.

Table F-1. Model for Predicting the Relative Citation Index Measured Using Whole Counts in the Expanding Journal Set

Variable (in order of entry)	Cumulative r-squared	Coefficient	Standard Error
Postdoctorates (per 100)	0.365	0.21	0.02
Type of Control – Public	0.466	-0.95	0.07
1994 Carnegie Class (Med)	0.532	1.56	0.11
1994 Carnegie Class (R-1)	0.576	1.26	0.08
State/Local Government Financed Academic R&D Expenditures (per $1M)	0.599	-0.028	0.003
Other Funding Sources for Academic R&D Expenditures (per $1M)	0.622	0.0540	0.006
S&E Graduate Students (per 100)	0.640	-0.431	0.032

Time	0.655	0.089	0.010
1994 Carnegie Class (M-1)	0.670	-1.66	0.16
Federally Financed Academic R&D Expenditures (per $1M)	0.683	0.011	0.001
Sources: NSF-NIH Survey of Graduate Student and Postdoctorates in Science and Engineering and NSF Survey of Research and Development Expenditures at Colleges and Universities.			

For the most part, the variables that were included in the stepwise regression were the same as or highly correlated with those identified in the linear regression. Primary similarities and differences are as follows:

- Federally financed R&D expenditures and other financed R&D expenditures entered the stepwise regression as previously. Non-federally financed R&D expenditures did not enter the regression. Instead, two components of non-federally financed R&D expenditures (i.e., state/local government financed and other financed expenditures) entered. Expenditure variables with smaller contributions in the previous regression (i.e., total basic research expenditures and institution-funded research expenditures) were not included.
- S&E graduate students entered the stepwise regression in place of fall enrollment of graduate students. However, the regression coefficient for S&E graduate students was negative, which implies that the effect is related to other independent variables. Undergraduate enrollment was not included.
- Total postdoctorates entered the stepwise regression, but other postdoctorate variables with smaller contributions in the previous regressions (i.e., postdoctorates with and without M.D.s supported by federal traineeships) did not enter.
- No non-faculty research staff variables entered the stepwise regression.
- No faculty count variables entered the stepwise regression.
- No degree awarded variables entered the stepwise regression. This may be partially attributable to the correlation between graduate degrees awarded and number of S&E graduate students.
- The same three Carnegie classifications entered the stepwise regression.
- Public control entered the stepwise regression as previously.
- The time variable entered the stepwise regression. This may partially reflect the effect of new journals entering the journal set over time.

Since the NRC ratings were only available at the field-group level, they could not be included in the analysis of relative citation counts at the institution-level described above. To examine their influence we entered the NRC ratings into regressions where the dependent variable was the relative citation count at the field-group level (except for the

medical sciences group, which was not NRC rated). The independent variables were all of those used in the institution-level regressions plus the NRC ratings. For three of the field groups, the variable that was entered first in the regression was the 1993 NRC SQR ratings. For those field-groups, the proportion of variance accounted for by these ratings was as follows: 1) bio-life-agricultural sciences (57.9%), 2) computer science (17.5%), and 3) social sciences and psychology (24.0%). For the eng-math-physical sciences field-group, NRC ratings entered third in the regression, with an incremental r-squared of 3.8%; when entered by itself, NRC ratings accounted for 9.3% of the variance. This suggests that NRC ratings are strongly associated with and predictive of the number of citations per publication.