

**How much is a collaboration worth?
A calibrated bibliometric model**

J. Sylvan Katz and Diana Hicks

j.s.katz@sussex.ac.uk

ESRC Centre for Science,
Technology, Energy and Environment Policy
Science Policy Research Unit
University of Sussex
Falmer, Brighton, BN1 9RF, UK

published in

Proceedings on the Sixth Conference of The International Society for Scientometric and Informetric,
Jerusalem, Isreal, June 16-19, 1997, pp. 163- 175

How much is a collaboration worth? A calibrated bibliometric model

J. Sylvan Katz and Diana Hicks

ESRC Centre for Science, Technology, Energy and Environment Policy, Science Policy Research Unit,
University of Sussex, Falmer, Brighton, BN1 9RF, UK

Interest in collaboration is increasing in policy circles. There are numerous international and national programs to encourage collaboration, for example, between university and industry researchers. However, little is known about the way in which collaboration changes the impact of a research publication. This paper explores how the impact (average citations per paper) varies with different types of collaboration. A calibrated bibliometric model is derived that demonstrates that collaborating with an author from the home institution or another domestic institution increases the average impact by approximately 0.75 citations while collaborating with an author from a foreign institution increases the impact by about 1.6 citations.

I. Introduction

Over recent years, there has been increasing interest among researchers and within science policy circles about the notion of research collaboration. It is widely assumed that collaboration is 'a good thing' and that it should be encouraged (Katz and Martin, 1997). Numerous initiatives by the European Commission and national governments have launched with the aim of developing collaboration among researchers - bringing them together, for instance, in distributed centres of excellence. There have also been policies aimed at improving the links between science and technology - in particular, between university and industry. Most governments are increasing the level of international collaboration in the belief that this will bring about many benefits, for example, cost-savings and higher impact research.

In the light of the collaboration policy we ask two questions. What collaboration strategy has the largest impact (citations per paper)? Does it make a difference if we choose to collaborate with a colleague from our own institution, another domestic institution or a foreign institution? Current scientometric research has demonstrated that on average the impact of a paper as measured in citations per paper is partially related to the number of authors, institutions and countries participating in the paper. For example Narin (1990) showed that multiple institution papers are more highly cited than single institution papers and papers with a foreign collaborator are more highly cited than domestic papers.

In the bibliometric community citation analyses are used to measure the impact of a research paper (Narin, 1976; Royal Society, 1986). This is not easy because factors such as the age of the paper, number of citing years and field-to-field citing cultures affect citation patterns. Citation techniques have been extensively studied and the limitations are reasonably well understood. To the best of our knowledge even though the general effect of various types of collaboration have been explored, few, if any, investigators have calibrated how impact varies with various types of collaboration. This paper will demonstrate that it is possible to determine how the impact changes on average as the number of authors from the same institution, other domestic institutions or foreign institutions participating in a paper changes. This result provides support for policies seeking to foster certain kinds of collaboration. It may also provide insight for researchers seeking to maximise impact.

II The data source

Over the past three years we constructed the BESST (Bibliometric Evaluation of Sectoral Scientific Trends) database (Katz et al, 1995) to help us visualise and model the UK system of collaborating institutions (Hicks and Katz, 1997; Katz and Hicks, 1996). This database contains unified bibliometric information on nearly a half million refereed UK R&D publications (articles, notes and reviews) indexed in the Science Citation Index (SCI) between 1981 and 1994. The SCI data with matched citations was purchased from the Institute of Scientific Information (ISI).

It is important to note that self-citations (i.e. citations by an author to his/her previous work) are included in the citation data. It is too costly in time and computational complexity to remove self-citations from 500,000 publications. Furthermore, we believe that not all self-citations should be removed, since it is common practice for researchers to build on their previous work and by necessity cite it. Perhaps, only excessive self-citations should be removed.

In order to calibrate the impact of various collaboration types we extracted just five items per paper from the BESST data: journal name, number of authors, number of UK institutions participating in the paper, the number of countries in which institutions resided, and the total number of citations received by the paper in the publication year plus the subsequent three years. The shorter four year citation window was used for policy purposes to explore citation patterns to more rapidly moving research.

ISI indexes the corporate addresses listed on each paper as published in a journal. Because of spelling errors or author inaccuracies when recording institutional names, the same institutional name may appear in a variety of forms. Also, it is not unusual for authors at the same institution or an author with an interdepartmental appointment to report each department separately, resulting in a duplicate listing of an institution's name. In order to overcome these difficulties we standardised the names of all UK institutions (Katz 1996). Furthermore, each standard name was assigned to one of seven institutional sectors: education, medical, research council, industry, government and non-profit.

In order to examine UK collaboration characteristics by scientific field we developed a science discipline/field classification based on ISI's subfield categories. In the ISI scheme a journal can be classified into one or more of about 150 scientific subfields. We use the subfield overlap to capture more of the inter-field and inter-disciplinary nature of science. This appears to be absent from most journal classification systems. We do not believe our classification system is flawless nor do we completely accept the discipline/field metaphor of science. However, we think that for general policy purposes the SPRU scheme is intuitive and functional (Katz and Hicks, 1995). The disciplines and subfields are listed in Table 1.

Table 1 - SPRU Discipline/Field Classification

<p>Life Sciences Discipline biological sciences agricultural sciences medical and health sciences inter-field life science</p>	<p>Natural Sciences Discipline physical sciences chemical sciences earth sciences mathematical sciences inter-field natural</p>
<p>Engineering and Materials Science Discipline engineering sciences materials science information, computers and communications technology inter-field engineering & materials science</p>	<p>Interdisciplinary Sciences Discipline life-natural sciences life-engineering & materials science natural-engineering & materials sciences life-natural-engineering & materials sciences</p>

The data for this paper are aggregated to the discipline level. We believe the data to be relatively noise-free and we have measured the error to be less than 2% . The analyses in this paper use a citation distribution database derived from the data we extracted from the BESST database. It will be described later.

III Collaboration type and citations per paper

As previously mentioned, it has been demonstrated that on average the number of citations a paper receives is dependent on factors such as the number of authors, institutions, countries participating in the paper and science field. Before we develop our calibration model we will demonstrate that the UK data are consistent

with Narin's results. Also we will demonstrate that in general UK trends are similar across disciplines and sectors.

Authors and countries

We determined how many single and multiple author papers with and without foreign collaborators were published between 1981 and 1991. We counted the cumulative number of citations to each type of paper and calculated the average impact (citations per paper). Table 2 summarises the findings.

Table 2 - Impact of author and foreign collaboration

Science	Authors	Countries	Papers (P)	Citations (C)	Ave. C/P
All	One and many	Any	376,226	2,411,789	6.4
	One	One ¹	74,481	285,536	3.8
	One	Many ²	1,505	7,705	5.1
	Many	One	236,592	1,525,400	6.5
	Many	Many	63,648	593,148	9.3
Life	One and many	Any	226,918	1,563,331	6.9
	One	One	38,817	168,492	4.3
	One	Many	487	2,776	5.7
	Many	One	156,693	1,072,777	6.9
	Many	Many	30,921	319,286	10.3
Natural	One and many	Any	90,506	497,009	5.5
	One	One	19,113	70,671	3.7
	One	Many	743	3,780	5.1
	Many	One	48,025	255,135	5.3
	Many	Many	22,625	167,423	7.4
Engineering & materials	One and many	Any	21,714	45,299	2.1
	One	One	6,186	8,939	1.5
	One	Many	77	184	2.4
	Many	One	12,647	28,980	2.3
	Many	Many	2,804	7,196	2.6
Multidisciplinary	One and many	Any	35,675	302,229	8.5
	One	One	9,785	36,172	3.7
	One	Many	183	915	5.0
	Many	One	18,594	166,721	9.0
	Many	Many	7,113	98,421	13.8

Notes:

1. One country means all institutional addresses were in the UK.
2. 'One author - many country' papers are those that list one author and multiple institutional addresses, one of which is not in the UK. This suggests the author had an affiliation with a domestic and a foreign institution. We interpret this as two institutions collaborating by 'sharing' a researcher.
3. The sum of the papers and citations across the individual disciplines does not equal the total number of papers published between 1981 and 1991. Approximately 0.5% of papers could not be classified into a discipline due to missing journal information at ISI.

We see that in every discipline collaborative papers have more impact than sole author papers. 'Many author-many country' papers have the highest average impact and 'one author - one country' the lowest impact of the five categories analysed. The highest impact papers are in the multidisciplinary sciences followed in decreasing order of impact by life, natural and engineering and materials sciences.

Authors and domestic institutions

In order to see if the same trend occurs at the national level we explored the relationship between the number of authors and the number of domestic (UK) institutions. Table 3 lists the number of single and multiple author papers involving one or many domestic institutions. Papers involving a foreign institution were not included so as not obscure the effect of domestic collaboration.

Table 3 - Impact of author and domestic institution collaboration

Science	Authors - domestic Institutions	Papers (P)	Citations (C)	Ave. C/P
All	One - One	72,350	278,514	3.9
	One-Many	2,131	7,022	3.3
	Many - One	175,741	1,086,179	6.2
	Many - Many	60,851	439,221	7.2
Life	One - One	37,100	162,741	4.4
	One-Many	1,717	5,751	3.3
	Many - One	112,933	74,6516	6.6
	Many - Many	43760	326261	7.5
Natural	One - One	18,926	69,951	3.7
	One-Many	185	716	3.9
	Many - One	37,967	197,715	5.2
	Many - Many	10,058	57,420	5.7
Engineering & materials	One - One	6,121	8,858	1.5
	One-Many	64	81	1.3
	Many - One	10,255	23,455	2.3
	Many - Many	2,392	5,525	2.3
Multidisciplinary	One - One	9,635	35,718	3.7
	One-Many	150	454	3.0
	Many - One	14,131	117,270	8.3
	Many - Many	4,463	49,451	11.1

Again we see that papers with domestic collaborations have more impact on average than single author papers and that in general the highest impact came from publications involving many authors at many institutions. In every discipline except the natural sciences, the lowest impact papers were published by one author giving two or more domestic institutional addresses on the paper. Also, we see that impact varies across science disciplines in the same way as it did in the previous case.

Domestic and foreign institutions

Next we explored the relationship between domestic and foreign institutional collaboration. Table 4 gives the impact for papers with various types of institutional collaboration: no collaboration, at least one domestic institution, at least one foreign institution, and at least two domestic and one foreign institution.

Table 4 - Impact of Domestic and foreign institutional collaboration

Science	Institutional Collaboration	Papers (P)	Citations (C)	Ave. C/P
All	None	248,091	1,364,693	5.5
	Domestic only	62,982	446,243	7.1
	Foreign only	55,749	481,704	8.6
	Domestic & foreign	9,404	119,149	12.7
Life	None	150,033	909,257	6.1
	Domestic only	45,477	332,012	7.3
	Foreign only	26,227	256,863	9.8
	Domestic & foreign	5,181	65,199	12.6
Natural	None	56,893	267,666	4.7
	Domestic only	10,245	58,140	5.7
	Foreign only	20,307	137,380	6.8
	Domestic & foreign	3,061	33,823	11.1
Engineering & materials	None	16,376	32,313	2.0
	Domestic only	2,457	5,606	2.3
	Foreign only	2,660	6,765	2.5
	Domestic & foreign	221	615	2.8
Multidisciplinary	None	23,766	152,988	6.4
	Domestic only	4,613	49,905	10.8
	Foreign only	6,388	79,963	12.5
	Domestic & foreign	908	19,373	21.3

The impact is highest for papers involving at least two domestic and one foreign institution followed in each type by foreign only and domestic only. Papers with no institutional collaboration had the lowest impact. Just as in the first case the highest impact comes from papers involving a foreign institution. Again, we see the impact varies across disciplines in a manner similar to the previous two cases.

Domestic and foreign institutions by sector

Now let's look at how impact varies with domestic and foreign collaboration by UK sector. Table 5 gives the impact for publications involving at least one industrial partner. Industry participates in 8% of the UK's refereed R&D publications. We chose this sector because with few exceptions it represents all the other UK sectors.

First, we see that when industry collaborates its papers have greater impact than when it does not collaborate. As in the previous case, except for the multidisciplinary sciences, the highest impact came from publications with a domestic and foreign collaboration. In general, the highest impact came from papers involving a foreign institution. Using industry papers as an example we demonstrate that the impact varies with collaboration type and across disciplines in a manner similar to that shown in the previous examples

Based on the preceding tables we see that generally speaking

1. collaborative papers have more impact than non-collaborative papers,
2. the highest impact papers involve a foreign institution,
3. characteristics 1 and 2 are broadly the same for each discipline and UK sector.

Table 5 - Impact of domestic and foreign collaboration with industry

Science	Institutional Collaboration	Papers (P)	Citations (C)	Ave. C/P
All	None	16,699	81,831	4.9
	Domestic only	10,188	63,021	6.2
	Foreign only	2,863	24,770	8.7
	Domestic & foreign	1,080	9,796	9.1
Life	None	6,091	45,274	7.4
	Domestic only	4,420	35,998	8.1
	Foreign only	1,283	13,303	10.4
	Domestic & foreign	488	5,306	10.9
Natural	None	4,108	15,377	3.7
	Domestic only	3,254	16,654	5.1
	Foreign only	763	5,624	7.4
	Domestic & foreign	349	2,767	7.9
Engineering & materials	None	3,905	7,737	2.0
	Domestic only	1,287	2,974	2.3
	Foreign only	364	907	2.5
	Domestic & foreign	110	378	3.4
Multidisciplinary	None	2,538	13,400	5.3
	Domestic only	1,200	7,351	6.1
	Foreign only	447	4,935	11.0
	Domestic & foreign	128	1,344	10.5

Observations one and two confirm Narin's (1990) findings but observation three suggests that, although the absolute value of the impact differs across sectors and disciplines the trend for the impact to increase with collaboration, especially with a foreign institution, is the same within sectors and disciplines. This suggests that collaborating with colleagues from the same institution, other domestic institutions and a foreign institution may lead to higher impact irrespective of sector or discipline. The question that remains unanswered is "can we disentangle the effects of adding an author, institution or country to calculate precisely the average effect each might be expected to have on the impact of a paper?"

IV. Calibrating alternative collaboration types

The calibration data are composed of the distribution of citations per paper by number of collaborating authors, institutions and countries. In other words we constructed a data set that contains the number of papers and citations to papers published by all combinations of the number of authors, UK institutions and countries. The data cannot be broken down by discipline or sector because there would be too few papers in each category to draw reliable conclusions. Table 6 lists three representative examples from the almost 3,000 combinations in the data set.

Table 6 - Sample citation distribution data

Number Authors	UK institutions	Number Countries	Number Papers	Number Citations	Citations per Paper
1	1	1	72350	72350	3.85
1	2	1	2014	6802	3.38
1	1	2	2831	14484	5.11

To obtain calibration curves we need to isolate the effect of adding another domestic institution from that of adding, for example, an author in the same institution. We will do this by isolating 'pure' cases from 'impure' cases. For example, if we want to examine 'how increasing numbers of authors within a domestic institution affects the impact' we examine papers that list one domestic institutional address (pure) and ignore papers with two or more domestic or foreign institutions (impure). Thus, using Boolean logic we isolated 'pure' cases from the data and examined how the impact changes as the number of (1) authors on a paper increase in a domestic institution, (2) collaborating domestic institutions increase; and (3) foreign institutions increase.

Authors from the same institution

We isolated publications with authors from the same institution by selecting papers with only one domestic institution and which do not involve a foreign country. Table 7 gives data showing how the impact changes as the number of authors from the same institution increases. Note, that only papers listing ten or less authors were used since these account for 99.9% of single domestic institution publications with no foreign collaborators.

Table 7 - Authors from the same institution

Authors	Papers (P)	Citations (C)	Ave. C/P
1	72,350	278,514	3.8
2	85,486	444,010	5.2
3	49,751	304,714	6.1
4	23,263	168,804	7.3
5	10,048	86,576	8.6
6	4,217	42,285	10.0
7	1,718	21,245	12.4
8	663	8,708	13.1
9	283	4,600	16.3
10	132	2,160	16.4

A regression analysis of citations per paper with number of authors yielded a relationship where the impact, I, increased exponentially with the number of authors, A, given as follows ($p < .001$).

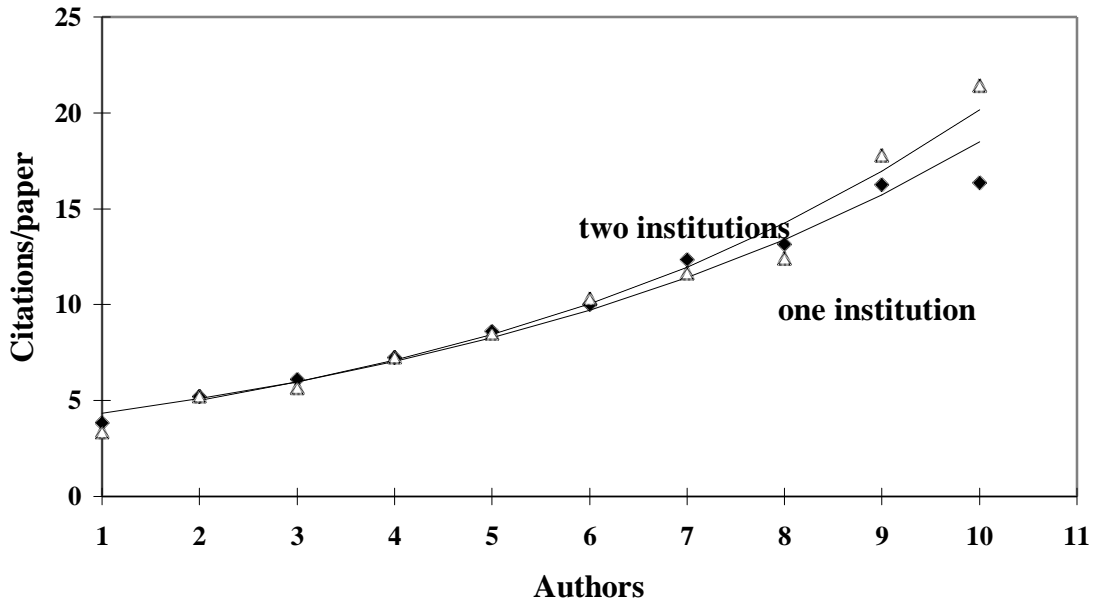
$$I = e^{(1.31 \pm 0.05 + 0.16 \pm 0.01 A)} \quad (1)$$

In order to check the validity of this relationship we explored one more case. We looked to see how impact changed as the number of authors increased on papers listing only two domestic institutions. This is not quite the same as the previous case because a given author could have resided at one or the other institution. The next equation ($p < .001$) gives the exponential relationship between citations per paper and number of authors in this case.

$$I = e^{(1.26 \pm 0.06 + 0.17 \pm 0.01 A)} \quad (2)$$

We can see that the slope and intercept of this equation are similar to the previous equation. Both trends are plotted in Figure 1. Except for what appears to be a small contribution from the additional institution when there are more than four authors at two institutions both lines indicate that the impact increases exponentially with the number of authors. The addition of an author from the same institution earns an additional 0.76 citations per paper while the addition of two authors earns an additional 1.64 citations per paper.

Figure 1 - Impact for collaborations with authors at the same institution



Domestic institutions

Using only papers that have the same number of authors as domestic institutions and no foreign countries we isolated papers with one author residing at each domestic institution and explored how impact changes as the number of domestic institutions increase. There are about 86,000 papers of this type. Table 8 gives the impact for papers with up to four domestic institutions and no foreign institutions. This accounted for 99.9% of the publications. Table 8 summarises the findings.

Table 8 - Collaboration with other domestic institutions

Authors	Domestic Institutions	Papers (P)	Citations (C)	Ave. C/P
1	1	72,350	27,8514	3.0
2	2	12,766	66,991	5.3
3	3	1,932	9,702	5.0
4	4	330	2,161	6.6

Even though there are only four data points we did a linear regression of citations per paper on number of domestic institutions, N, and found that impact increases linearly with the number of domestic institutions involved in the research publication. The relationship is given in the following equation ($p < 0.05$)

$$I = 3.25 \pm 0.70 + 0.78 \pm 0.26 N \quad (3)$$

It seems that each author from an additional domestic institution earns 0.78 citations per paper. Thus adding an author from another domestic institution has about the same effect as adding an author from within the same institution. However, if two authors (i.e. three in total) from two other domestic institutions are added they earn an additional 1.56 citations per paper. This is about the same impact as collaborating with two authors from the home institutions. Four institution papers account for 99.9% of institutional papers with more than four institutions.

Foreign institutions

Using papers listing only one domestic institution and as many authors as countries we can isolate the effect of adding an author from a foreign institution. Table 9 gives the impact as the number of foreign countries increase and accounts for 99% of papers with this type of collaboration.

Table 9 - Collaborations with foreign institutions

Authors	Countries	Papers (P)	Citations (C)	Ave. C/P
1	1	72,350	278,514	3.9
2	2	14,812	86,082	5.8
3	3	1,394	9,839	7.1
4	4	151	1,334	8.8

Again a regression analysis yielded a linear relationship between impact and the number of foreign institutions, F, listed on the paper. This relationship is expressed in the following equation ($p < .001$)

$$I = 2.40 \pm 0.21 + 1.60 \pm 0.08 F \quad (4)$$

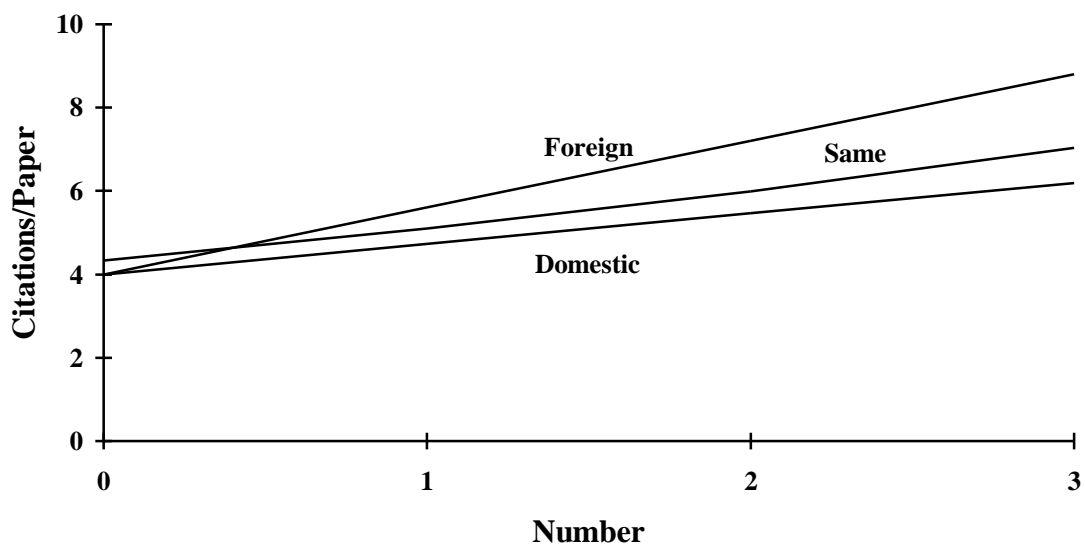
Adding an author from a foreign institution to a paper earns an additional 1.60 citations per paper on average which is more than twice as much as adding an author from a domestic institution.

Summary of citations per paper by collaboration type

Figure 2 combines the results from the three different cases in one graph: same institution, domestic institution and foreign institution. We only provide the regression curves for papers with up to three authors, institutions or countries since this accounts for the majority of papers published by UK authors.

The calibrated bibliometric model of collaboration suggests that adding one more author from the same institution or another domestic institution earns approximately an additional 0.75 citations and from a foreign institution 1.6 citations. Similarly, adding two more authors from the same institution or another domestic institution earns an additional approximately 1.6 citations and from two foreign institutions 3.2 citations.

Figure 2 - Overall calibration curve of impact by collaboration type



The calibration curves were derived for papers from all disciplines and sectors. Unfortunately the BESST database is not large enough to do the same analysis for individual disciplines and sectors. However, if the trends shown in the first section of this paper hold we might expect similar trends but with the calibration curves having different slopes and intercepts reflecting the unique characteristics of the individual disciplines and sectors. In order to test this hypothesis, we estimate we would need about ten times as many papers in our database. In the future we hope to investigate the change in impact for the mixed collaboration type.

V. Conclusions

First, we have demonstrated that, in general, the impact of UK papers in any discipline or sector are higher if there is a collaboration of some kind. Sole author papers appear to be in the minority and have less impact. The highest impact comes from publications involving collaboration with foreign institutions. Second, using the distribution of impact by the number of authors, domestic institutions and countries we can calibrate how impact increases as the number of same, domestic and foreign authors increase. The impact increases exponentially with the number of collaborating authors from the same institution. Impact increases linearly with increasing number of domestic and foreign institutions. However, papers involving collaboration with a foreign institution have greater impact than papers with collaborations with an author from the same or another domestic institution.

References

- Hicks D and Katz JS (1997) *The changing shape of British Industrial Research*, STEEP special report No 6, SPRU, University of Sussex, UK.
- Hicks DM and Katz JS (1996) Hospitals: a hidden research system, *Science and Public Policy*, 23 (5).
- Katz JS, Hicks DM, Sharp M & Martin BR, (1995) *The Changing Shape of British Science*, STEEP report No 3, SPRU, University of Sussex, UK.
- Katz JS and Hicks D (1995) The Classification of Interdisciplinary Journals: A new approach. *Proceedings of the Fifth Biennial Conference of the International Society for Scientometrics and Informatics* (River Forest, Ill, USA).
- Katz JS and Hicks DM (1996) A Systemic View of British Science, *Scientometrics*, 35 (1) 133-154
- Katz JS and Martin BR, (1997) What is Research Collaboration?, *Research Policy*. 26 (1)
- Narin F (1976) *Evaluative Bibliometrics: The Use of Publication and Citation Analysis in the Evaluation of Scientific Activity*, Contract NSF C-627, National Science Foundation. Monograph: 456 pp. NTIS Accession #PB252339/AS.
- Narin F and Whitlow ES (1990) "Measurement of Scientific Co-operation and Coauthorship in CEC-related areas of Science", Report EUR 12900, Office for Official Publications in the European Communities, Luxembourg.
- The Royal Society (1986) *Evaluation of National Performance in Basic Research*. Advisory Board for the Research Councils Policy Studies No, Department of Education and Science (London, UK)

Acknowledgements

The BESST project was funded by the UK Office of Science and Technology, Department of Trade and Industry, Department of Health, Medical Research Council, Economic and Social Sciences Research Council and Engineering and Physical Sciences Research Council