

# Studies in Scientometrics II

## The Relation between Source Author and Cited Author Populations

Derek de Solla Price  
Suha Gürsey  
Yale University

### ABSTRACT

*A sequence of annual Citation Indexes behaves in very much the same way as the sequence of Source Indexes. Five years of Source Index were regarded (1964-1968) as sufficient to distinguish the transient and continuant classes amongst the authors, and three years of the Citation Index (1966-1968) to pick up references to authors commencing publication at the beginning of the period. The file listed 640 authors of which 381 were in the five Source Indexes and 442 in the three Citation Indexes with 183 names being common to both sets of index. Of the transient authors who appear in only a single Source Index 71% are uncited, 19% are transiently cited and only 10% seem to be continuants in citation. A transient author has about 30% chance of being cited in any year. Of the core continuant authors who publish in all five years of record, 19 of the 22 are also cited in all the three years. It's worth noting that these 19 authors are also those who continue not merely for a five year span but for the seven years recorded in the previous study. Continuant source authors have a strong tendency to be continuants also in citation. A considerable group of 259 names in Citation Index do not appear amongst the source authors. Of the 259 names 80% are transiently cited and 20% are continuants. With the addition of new citation data and the characteristics of the non-source groups, not before detected, the demographic groups now become as: a) transients b) non-core continuants c) core continuants d) terminated continuants e) former transients f) newly recruited continuants.*

Recently we have used the technique of taking a small alphabetic slice of a sequence of several annual Source Indexes of the Science Citation Index to elucidate the demographic structure of the scientific and technical publishing population.\*

This technique is now extended to cover a series of Citation Indexes of the same volumes of the S.C.I. so as to determine the relation between the demographic groups of authors and the extent to which they are drawn on by the rest of the scientific community. It also reveals some information about the class of scientists who are cited but do not appear amongst the indexes of current authors. First it must be remarked that the Citation Indexes of all the references made in this year's papers are necessarily much larger and more extensive than the corresponding Source Indexes. Usually they contain about 1.5 times as many names for although they are recorded only by the name of the first author of the cited paper, the references in papers of this year extend back through the entire time span of previous publication and not only to all published journals without anyone making a selective choice, but also to non-journal items such as theses, printed books, patents, the report literature and even to informal publication when this is specifically acknowledged. Even the most cursory examination is sufficient to reveal that a sequence of annual Citation Indexes behaves in very much the same way as the sequence of Source Indexes which has already been subjected to analysis. Two consecutive Citation Indexes overlap so that about half the names are carried over directly to the next year, and an additional quota of names are carried over and reappear after a gap of a year

\* See: Derek de Solla Price and Suha Gürsey: Studies in Scientometrics I: Transience and Continuance in Scientific Authorship. *Ciência da Informação*, 4(1):27-40, 1975. We should also like to express our deep gratitude to Dr. E. Garfield and the officers and staff of I.S.I. for their very effective long standing cooperation and technical help in this and related projects. The investigation was supported under Grant GS-39830X from the National Science Foundation.

or more. Each year there is a new supply of names that have never before been cited, and undoubtedly an investigation would reveal that there are also a set of names that never occur again. Thus we may suppose that there will exist classes of transients, and continuants, among the cited authors too, and further there will be authors cited so frequently that they must be counted as core continuants because of their appearance in every annual Citation Index over a considerable period. The parameters of probability and frequency governing this structure may be somewhat different for cited rather than source authors, but to a first approximation it would seem that the similarity is curiously close. Using the same alphabetic sample as in the previous examination a careful search was made for all names in three consecutive Citation Indexes, all names being edited into the same comprehensive list. Table I shows the empirical data derived for a span of Source Indexes 1964 through 1968, and Citation Indexes 1966 through 1968. Five years of Source Index were regarded as sufficient to distinguish the transient and continuant classes amongst the authors, and the Citation Index was begun a couple of years later to pick up references to authors commencing publication at the beginning of the period. Even this short run is quite sufficient to show the main features. Basically our file listed 640 authors of which 381 were in the five Source Indexes and 442 in the three Citation Indexes, with 183 names being common to both sets of index. It is worth noting that as in the case of each Index separately, about half the names in the Source Index set are "carried forward" into the Citation Index set. Indeed the intersection between any two annual indexes, Source or Citation, not too far distant from each other in time, seems to be of size about half the smaller of the pair of indexes. From Table I it is clear that there are strong concentrations in two groups. A large number of authors appear in only a single Source Index and in none of the Citation Indexes. Another large group appear in all the Source Indexes and all the Citation Indexes. More particularly, of the transient authors who appear in only a single Source Index, 71% are uncited, a further 19% are transiently cited in a single Index, and only 10% seem to be continuants in citation. These data are in good agreement with the simple probability model in which the work of a transient author has about a 30% chance of being cited in any year; on that assumption there would be 70% authors uncited, 21% cited just one year, 6% cited for two years, 1% for three years, 1% for more than three years. For the core continuant authors who publish in all five years of

record, 19 of the 22 are also cited in all three years and the other 3 are cited in two of the three years. It is worth noting that the 19 who are core continuants in both categories are also those who continue not merely for a five year span but for the seven years recorded in the previous study. For the non-core continuants of two, three and four years as a source, the incidence of appearance as a cited author is intermediate in behavior between that of transients and that of the core. For the group as a whole the behavior is precisely the opposite of the transients in that they have about a 70% of being cited, 30% uncited. As the number of years of continuance increases so does the chance of citation as shown in Table II. The data probably overestimate the numbers of non-core continuants who are uncited, for as can be seen from Table I there are 15 cases of authors publishing in 1967/68 and 5 cases of those publishing in the same years and a previous year as well who were possibly uncited because their publications were still too recent. It follows that continuant source authors have a strong tendency to be continuants also in citation. In addition to the demographic groups already discussed we have in the Citation Indexes a considerable group of people who were cited in one or more of the annual indexes but whose names do not appear amongst the source authors. Of the 259 names 80% are transiently cited in a single year, 15% appear in two years, and only 5% in all three Citation Indexes. We must, it seems, be dealing here with a mix of at least two different groups of authors. On the one hand we have a group of former transient source authors and on the other the relatively small number of authors who were once continuants and are still fairly heavily cited even though they have retired from publication or died. The number of such persistently cited people not appearing in the Source Indexes amounts to 52 compared with an active continuant population of about 160 people in this period. This implies that about 75% of all the continuants that have ever been are still writing, but unquestionably the survival rate is higher than this for some of the people cited must have published only in books or in journals not on the Source List of I.S.I. The number that have ceased is thus less than 25% of the number of continuants that have ever been; this means that the number of continuants must have grown through more than a factor of 4 — at least two doubling periods — in the interval of time it takes an average continuant to retire. With a growth rate of the usual value of 7 - 10% per annum this gives some 15 — 20 years as the publishing life span of a continuant.

Such a life span seems rather too long for an average value between the few who spend a full working life at the research front and the many who spend one or two three-year terms in junior appointment. The forgoing calculation needs modification if we suppose, as seems reasonable, that the 52 non-source authors that appear in two or three Citation Indexes are not the only retired continuants. Suppose that there were in fact a retired continuant population of size  $C$ , each of which had a probability  $p$  of being in an annual Citation Index, and that successive appearances were quite independent. Then the number of cases of appearance in at least two such successive lists is given by  $Cp^2$ , and the number of cases of at least three appearances is  $Cp^3$ . Table I shows these numbers to be about 25 and 12 respectively, hence  $p$  is approximately  $1/2$  and  $C = 100$ . There are then about 100 terminated continuants instead of 52 and the number of them cited just in one index out of three is expected to be  $100 (\frac{1}{2})^3 = 12$ . These 12 must be subtracted now from these who were previously counted amongst the transiently cited former transients.

We thus derive the new result that one component of former authors who are now cited consists of about 57 authors per year picked out of the very large number of former transient source authors and cited just once in a single year. The other component consists of about 100 former continuants who are no longer source authors but who have a chance of about one half of being cited in any particular year. The number of terminated continuants being 100, and the number active being ca. 160 as before, the ratio of total born over total died is 2.6 and this implies that with a doubling period of 10 years we must take the publication lifetime of a continuant as about 14 years. If one were to assume that all of these 259 cited non-source authors were terminated continuants, and none of them former transients, the ratio would be 1.6 and the average lifetime a little less than 7 years. If one were to suppose that there are even more former continuants who happen to be uncited in the three indexes, the ratio and therefore the lifetime would be correspondingly reduced. In order to make the lifetime as short as half a doubling period, say 5 years, we would have to suppose the existence of 384 terminated continuants, so that the 259 actually found in the three years of Citation Indexes would be only  $2/3$  of those actually existing. After these considerations we are able to restate the categories of the scientific publishing community, amplifying the results previously found by the addition of the new citation data and the characteristics of the non-source groups not before detected. The demographic groups now become as

follows:

- a) Transients whose names occur only in a single Source Index. They usually have but a single authorship (often secondary) in this year, and only a small fraction of these authorships, perhaps a quarter are thereafter cited, usually just once in any year.
- b) **Non-core** continuants whose names occur in several successive Source Indexes. They have two or more authorships per year, and have about a 70% chance of being cited in any year.
- c) Core continuants whose names appear unfaillingly in a long sequence of Source Indexes. They are about the most prolific 20% of all continuants, have four or more authorships per year, and also unfaillingly appear in the Citation Indexes.
- d) Terminated continuants who formerly published for several years but remain cited with a probability of about 50% of appearing in any annual Citation Index.
- e) Former transients from all previous years must be a population about twice as numerous as the continuant community existing at any particular time. Though they are therefore about 67% of the scientific population, they have published less than 25% of all papers, and these papers probably account for less than 10%, perhaps less than 5% of all citations.
- f) **Newly recruited** continuants cannot in their first year be distinguished from transients.

Finally, as an additional bonus from the methodology of this investigation we are able to produce for this small alphabetic slice of the scientific population a pair of tabulations of the core groups who are continuant in the Source and Citation Indexes. Table III shows first the most cited authors during the three year period, together with their output of source authorship during the five year period, next it gives the most prolific authors with their citation record during the same period. There are 26 highly cited authors with an average of 10 or more citations/year, including two who were not listed in the Source Indexes, and there are 25 prolific authors with an average of 2 or more papers/year counting all primary and secondary authorships as a unit. Common to both lists are some 15 authors so that the overlap is about 60% of either, a very striking correlation when one remembers that these are but the top sections of a pair of lists having 640 names each and only 147 in common. It would seem that the assumption that there exists a highly cited and highly prolific core group is quite justified. It suggests further that highly placed names not common to the two lists merit special

investigation before letting these scores act as an evaluation. We have two clear cases of terminated continuants, several cases of the prolific production of apparently little cited work, and several cases of highly cited authors of rather few papers. In the last class however it must be remembered that some of these highly cited authors might be in the course of a termination and producing much less than in the period for which they are cited. In only one case is there something as flagrant as an author who has published 10 papers but received no citations.

#### RESUMO

*Uma seqüência anual de Citation Indexes mantém um padrão de comportamento em muito semelhante a uma seqüência de Source Indexes. Cinco anos do Source Index (1964-1968) foram considerados como suficientes para distinguir, entre os autores, as classes de transientes e permanentes, e três anos do Citation Index (1966-1968) para identificar referências a autores iniciando publicação no começo do período. O arquivo listou 640 autores, dos quais 381 estavam*

*incluídos nos cinco Source Indexes e 442 nos três Citation Indexes, sendo 183 nomes comuns a ambas categorias de índices. Dos autores transientes que aparecem apenas em um único Source Index, 71% não são citados, 19% são transitoriamente citados e somente 10% parecem ser permanentes em citação. Um autor transiente tem cerca de 30% de chance de ser citado em qualquer ano. Do núcleo de 22 autores permanentes que publicam nos cinco anos registrados, 19 são também citados em todos os três anos. É importante notar que esses mesmos 19 autores são também aqueles que continuam não apenas por um período de 5 anos mas pelo período de 7 anos registrados no estudo anterior. Os autores permanentes do Source tendem a ser permanentes também em citação. Um grupo de 259 autores que são citados não aparece entre os autores-fonte. Desses, 80% são transientes e 20% são permanentes. Com a edição de novos dados de citação e de características de grupos de autores não incluídos no Source, os grupos demográficos são distribuídos nas seguintes categorias: a) transientes, b) permanentes que não fazem parte do núcleo, c) permanentes integrantes do núcleo, d) permanentes que continuam a ser citados embora não mais publiquem, e) antigos transientes e f) novos permanentes.*

STUDIES IN SCIENTOMETRICS II

TABLE I

cited auth. source auth.	Not Cited	66	67	68	66/7	66/8	67/8	66/7/8	Total
	Not Source	Not Applicable	58	79	70	14	13	13	12
64	20		1		1		2	2	26
65	30	3					1	1	35
66	26	6	4	4		3	2	3	48
67	37	3	5	5	1	1	2	1	55
68	43	1	2	9		1		4	60
64/65	4						1	1	6
64/66	1	1				1	1		4
64/67	1		1	2			1		5
64/68							1	1	2
65/66	3	1		1		2		2	9
65/67	3	1					1		5
65/68				3					3
66/67	1	1	2		1		1	3	9
66/68	1			3		3		1	8
67/68	15			3			1	1	20
64/65/66	1						1		2
64/65/67	1						1	4	6
64/65/68									
64/66/67							1	1	2
64/66/68				1				2	3
64/67/68							1		1
65/66/67	1			1	1			1	4
65/66/68	1					1	2	1	5
65/67/68	2							2	4
66/67/68	3			2			2	7	14
64/65/66/67	1							3	4
64/65/66/68						1		2	3
64/65/67/68	1						1	2	4
64/66/67/68	1						1	3	5
65/66/67/68	1		1	1				3	7
64/65/66/67/68					1		2	19	22
Total	198	75	95	105	20	26	39	82	640

TABLE II

Years as source	Number of Years Cited				2
	0	1	2	3	
2	41	27	20	12	100%
3	22	10	24	44	100%
4	17	9	17	57	100%

DEREK DE SOLLA PRICE & SUHA GÜRSEY

TABLE III (\* authors common to both lists)

26 Most Cited Authors	Citations	Authorships		Total Papers	Years as Source
		Primary	Secondary		
*Palade GE	1441	6	37	43	5
*Palay SL	891	2	10	12	5
*Palatnik LS	392	51	21	72	5
*Pais A	262	14	12	26	5
*Paivio A	250	14	15	29	5
Pake GE	250	2	1	3	3
Pak WL	135	8	—	8	5
*Pacecek E	130	9	6	15	5
Pakula R	112	6	1	7	3
*Paladini AC	108	—	20	20	4
Pal MK	101	2	6	8	3
*Paladino AE	87	12	3	15	5
Pakiser LC	71	4	3	7	5
Pal Y	65	1	4	5	3
*Pal L	60	2	12	14	4
*Pakrashi SC	59	15	1	16	5
Pal S	51	7	2	9	4
Palais RS	48	2	—	2	2
*Pais M	44	7	5	12	4
*Pala G	39	15	10	25	5
Paldino RL	35	3	3	6	5
*Pakvasa S	32	9	9	18	5
*Palaic D	32	8	2	10	5
*Pakkenberg H	30	12	8	20	5
add to these, non-sources:					
Palache C	189	0	0	0	0
Pal BP	40	0	0	0	0

  

25 Most Prolific Authors	Citations	Authorship		Total Papers
		Primary	Secondary	
*Palatnik LS	392	51	21	72
*Palade GE	1441	6	37	43
*Paivio A	250	14	15	29
*Pais A	262	14	12	26
*Pala G	39	15	10	25
*Paladini AC	106	—	20	20
*Pakkenberg H	30	12	8	20
Paknikar SK	8	—	19	19
*Pakvasa S	32	9	9	18
Pakhomov VJ	18	6	10	16
*Pakrashi SC	59	15	1	16
*Palacecek E	130	9	6	15
*Paladino AE	87	12	3	15
Pak CYC	20	8	6	14
*Pal L	60	2	12	14
Pal AK	15	7	7	14
*Pais M	44	7	5	12
*Palay SL	891	2	10	12
Palacek J	3	7	5	12
Pakalns P	28	11	—	11
Pakala WE	6	5	7	12
*Palaic D	32	8	2	10
Palamidessi G	4	4	6	10
Palácios O	5	4	6	10
Pak MS	0	—	10	10