Scientometrics, Vol. 2, No. 1 (1980) 35–52

PUBLICATION RATE AS A FUNCTION OF LABORATORY SIZE IN A BIOMEDICAL RESEARCH INSTITUTION

J. E. COHEN

The Rockefeller University, 1230 York Avenue, New York, NY 10 021 (USA)

(Received February 20, 1979)

At the Rockefeller University in 1977–78, the number of all publications of a research group in a year was approximately proportional to the number of individuals in that group during the year. The number of primary research publications of a group in a year was also approximately proportional to the number of individuals in that group during the year. The observed frequency distribution of laboratory size was statistically indistinguishable from a 0-truncated negative binomial distribution, which is the equilibrium frequency distribution of size predicted by stochastic models for the dynamics of freely-forming primate social groups.

Introduction and summary

An analysis of information published for other purposes by the Rockefeller University indicates that, in the scientific disciplines represented there and under the administrative arrangements of that institution, the number of all publications of a research group in a year is approximately proportional to the number of individuals in that group during the year. The number of primary research publications ("primary" is defined in section Materials and methods) of a group in a year is also approximately proportional to the number of individuals in that group during the year. Thus, the publication rate per capita and the primary publication rate per capita are approximately independent of laboratory size.

The observed frequency distribution of laboratory size is statistically indistinguishable from a 0-truncated negative binomal distribution. This theoretical distribution is the equilibrium frequency distribution of size predicted by stochastic models for the dynamics of systems of freely-forming primate (human and nonhuman) social groups.

Materials and methods

The Rockefeller University is devoted promarily to research in the biological and biomedical sciences. There are substantial research activities in the behavioral and physical and mathematical sciences. There is no undergraduate program. The education offered to graduate students and postdoctoral fellows is primarily that of research apprenticeship. As graduate students routinely publish the results of their research, no distinction will be made between them and faculty in the following head-counts.

The University is suitable for investigating the effect of research group size on publication rate because "laboratories, rather than conventional categorical departments, are the fundamental units of the University. Each of its more than 60 laboratories typically includes a senior professor, several other faculty members, and postdoctoral fellows and graduate students who share the common scientific interests of the group".[1]

At the beginning of each academic year, the University publishes a catalog describing the subject matter of each laboratory and listing the academic members of the laboratory. Supporting staff are excluded. The membership of a laboratory may change during the academic year because of arrivals or departures. We take the number of individuals listed in the catalog as the number present.

For this study, we classify each laboratory listed in the 1977–78 catalog in one of five categories: (1) the behavioral sciences (8 laboratories); (2) biochemistry and cell biology (28 laboratories); (3) medicine and physiology, at the level of organs or higher (12 laboratories); (4) chemistry (inorganic, organic, and physical), physics, mathematics, and related sciences (12 laboratories); (5) history of science (2 laboratories).

The two "laboratories" in the history of science category each consist of single professors who are also heads of experimental laboratories. We do not consider this fifth category further here. Of the 60 remaining laboratories, some are directed jointly by two individuals. One professor also directed two laboratories listed separately. These exceptional situations receive no special treatment in the following analysis.

After each academic year, the Rockefeller University publishes an Annual Report, which lists the publications of each laboratory that appeared from 1 July to 30 June. Abstracts are excluded but book reviews or brief reports are not. A publication that is co-authored by members of two or more different laboratories is listed by each contributing laboratory. Here each title counts as one publication, regardless of length or number of co-authors.

For this study, a primary publication is identified by one of three kinds of citations: a publication with journal title, volume, and page numbers; sole or joint authorship of a book; or a doctoral thesis. Not counted as primary publications are chapters contributed to edited books or other collective works; book reviews; editorials in scientific or medical journals; and the editorship of collective works.

A single-author publication is identified by any citation listed with a single author, excluding editorship by a single person of a collective work.

For each laboratory, the category, number of people, number of all publications, number of primary publications, and number of single-author publications in the 1977–78 Annual Report were recorded in machine-readable form. Computations and graphical data analysis were carried out in APL on the time-sharing system of the City University of New York. In the computer-generated scatter diagrams, a printed point may indicate a coincidence of more than one laboratory.

The major inferences made here depend only on graphic presentation of the data, not on probabilistic assumptions of classical statistical analysis. However, certain results are asserted to be statistically significant or not significant. Since the data represent an exhaustive enumeration of the laboratories at Rockefeller University during 1977–78, not a random sample, it is necessary to consider in what sense the assumptions of conventional statistical tests are appropriate.

For example, the mean size of biochemistry and cell biology laboratories is 11.3 people while the mean size of laboratories in medicine and physiology is 10.8 people. This difference is real because all laboratories have been enumerated. We assert that the difference is not statistically significant because of the large variation in size within each category. We interpret this to mean that the observed frequency histograms of size estimate the probability density functions of size of a population of comparable laboratories. Viewed as samples from these populations of laboratories, the laboratories at Rockefeller University do not provide significant evidence of a difference in mean sizes between the populations.

For each category, we estimate the increase in number of publications resulting from an additional person in a laboratory in three ways:[2]

(1) by the slope coefficient in the fitted least squares line, assuming the variance in publications is independent of laboratory size;

(2) by the slope coefficient in the fitted least squares line through the origin, assuming the variance in publications proportional to laboratory size (the slope is then the ratio of the average, over all laboratories, of number of publications to the average, over all laboratories, of number of people);

(3) by the slope coefficient of the fitted least-squares straight line through the origin, assuming that the standard deviation of publications, rather than the vari-

ance, is proportional to laboratory size (the slope is obtained by computing, for each laboratory, the ratio of publications to people and then taking the average of that ratio over all laboratories in each category).

Results

For the benefit of other analysts, Table 1 presents the raw data.

Disregarding the extremely rare duplications between laboratories, there are 618 academic members listed for the 60 laboratories. Of the 631 total publications, 182 are by single authors and 474 are primary publications. Per laboratory, there is an average of 10.5 publications, of which an average of 3.0 are singly authored and an average of 7.9 are primary. The standard deviations of the numbers per laboratory of publications, single-author publications, and primary publications are, respectively, 8.8, 3.4 and 6.7.

The current fund expenditures and mandatory transfers for 1977–78 are \$38 015 000.[3] This figure includes all research and education plus overheads for support and administration. On this basis, the current expenditure per academic member of a laboratory is \$61 500. The current expenditure per publication is \$60 200 and per primary publication is \$80 200.

Before describing the relation between number of publications and laboratory size, we examine the distribution of laboratory size.

Laboratory size. The mean laboratory size is 10.30 with a standard deviation of 7.37 (range: 1 to 27). If the variation in laboratory size arose from purely random fluctuations in the number of individuals aggregated to form a laboratory, the size distribution would be given by the 0-truncated Poisson distribution. The truncated Poisson variance test[4] reveals that the observed frequency distribution is significantly overdispersed ($P < 10^{-4}$), that is, the observed variance is too large to have arisen from random sampling of a 0-truncated Poisson distribution.

This overdispersion might be due to differences among the categories in their mean laboratory sizes. The differences in mean laboratory size among categories are not significant overall according to one-way analysis of variance of either the head counts or of the square-roots of laboratory size, even through all 12 laboratories in chemistry, physics and mathematics (category 4) have not more than 12 people, which is the median size of laboratories in biochemistry and cell biology (category 2). Within each of the categories except that of chemistry, physics, and mathematics (category 4), laboratory size is significantly overdispersed ($P < 10^{-3}$ by the truncated Poisson variance test).

The frequency distribution of size of all laboratories is not significantly different from a 0-truncated negative binomial distribution (Table 2). Negative binomial

Table 1
Raw Data for 60 Laboratories at the Rockefeller University 1977–78:
(A) Subject matter category, (B) Number of academic people, (C) Number of publications, (D) Number of single-author publications, and (E) Number of primary publications

A	B	C	D	E	A	B	C	D	E
1	6	7	5	3	2	2	4	2	1
1	18	30	17	17	2	5	9	1	8
1	3	3	0	2	2	15	15	2	13
1	13	24	8	11	2	3	2	1	1
1	5	15	11	7	2	12	9	2	8
1	10	10	2	7	2	27	21	1	21
1	25	31	11	23	3	13	2	1	1
1	23	33	11	20	3	1	6	2	3
2	19	17	3	17	3	9	5	1	4
2	22	25	8	17	3	16	10	5	7
2	3	1	1	1	3	4	7	2	4
2	3	1	0	0	3	10	4	0	4
2	1	2	0	2	3	11	4	2	2
2	6	4	2	1	3	18	17	2	14
2	12	16	2	10	3	15	26	3	19
2	8	17	5	16	3	7	2	0	2
2	2	0	0	0	3	7	7	2	6
2	4	6	2	6	3	19	17	6	17
2	14	17	2	15	4	6	3	2	2
2	12	7	2	5	4	11	5	1	4
2	13	8	3	4	4	3	1	1	0
2	14	14	3	12	4	5	4	0	4
2	5	1	0	1	4	5	13	3	10
2	26	26	8	21	4	6	4	1	3
2	26	13	0	12	4	1	2	2	2
2	2	3	0	3	4	6	4	0	4
2	23	26	4	12	4	6	8	3	6
2	2	0	0	0	4	5	10	7	10
2	19	11	1	9	4	12	24	6	24
2	16	8	2	8	4	3	10	8	8

distributions are not fitted to the frequency distributions of size of the four categories separately because the number of laboratories in each category is too small to give a reasonable chance of rejecting the negative binomial model even if it were false.

Publications and laboratory size. A scatter diagram of the number of publications as a function of laboratory size for all 60 laboratories (Fig. 1A) offers no

Table 2
The observed frequency distribution of laboratory size and a fitted 0-truncated negative binomial distribution

Size	Frequency		Size	Frequency		Size	Frequency	
	Obs.	Pred.		Obs.	Pred.		Obs.	Pred.
1– 2	7	5.9	1– 3	13	9.7	1– 4	15	13.8
3– 4	8	7.9	4– 6	14	12.3	5– 8	15	15.9
5– 6	12	8.3	7– 9	4	11.1	9–12	9	12.1
7– 8	3	7.6	10–12	8	8.6	13–16	9	7.8
9–10	3	6.6	13–15	7	6.2	17–20	5	4.7
11–12	6	5.5	16–18	4	4.3	21–24	3	2.7
13–14	5	4.4	19–21	3	2.8	25–	4	3.0
15–16	4	3.4	22–24	3	1.8			
17–18	2	2.7	25–	4	3.0			
19–20	3	2.0						
21–22	1	1.5						
23–24	2	1.1						
25–26	3	0.8						
27–	1	2.2						
	$X^2 = 14.669$			$X^2 = 7.065$			$X^2 = 1.480$	
	df = 11			df = 6			df = 4	
	$0.1 < P < 0.2$			$0.3 < P < 0.4$			$0.8 < P < 0.9$	

The three displays above use the same observed distribution (Obs.) and predicted distribution (Pred.) with size classes pooled in three different ways. The parameters p = 0.18004, r = 2.20060 of the 0-truncated negative binomial distribution, probability that laboratory size is

$$k = \frac{1}{1 - p^r}\begin{bmatrix} r+k-1 \\ k \end{bmatrix} p^r(1-p)^k, \quad k = 1, 2, \ldots$$

were estimated using the method of *Brass* and the observed frequencies for each possible laboratory size prior to pooling. The moments of the observed frequency distribution satisfy the criteria of *Sampford* for a truncated negative binomial distribution. See *Cohen*[4] for details and references.

suggestion of a non-linear relationship. The least squares line, weighting each laboratory equally, is: publications = 0.939 + 0.923 X people. This line passes very nearly through the origin. A causal interpretation of this linear description is that an additional 10 people results in an additional 9 publications.

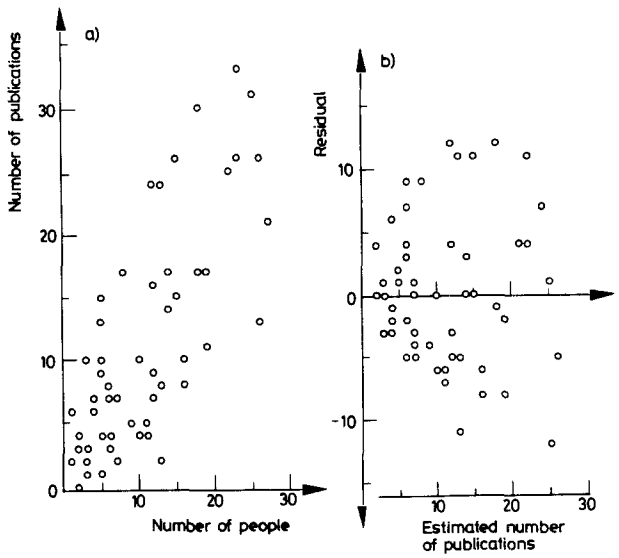

Fig. 1. For all laboratories, (A) the number of publications as a function of the number of individuals, and (B) the residual (difference between the observed number of publications and the number of publications estimated from a fitted least squares line) as a function of the number estimated

A more sensitive way of looking for deviations from linearity is to plot the residuals from this fitted line as a function of the estimated values.[5] Fig. 1B plots the observed number of publications minus the number of publications estimated from the straight line (the residual) on the vertical coordinate against the estimated number of publications on the horizontal coordinate. There is no suggestion of systematic concavity or convexity.

The variability of the number of publications increases as the size of laboratory increases. If it is supposed that the variance in number of publications is proportional to the size of laboratory, and that the true straight line must pass through the origin since laboratories of size 0 publish 0 papers, the least squares equation using method (2) is: publications = 1.015 X people.

The approximately linear relationship between publications and people when all laboratories are considered together could conceal marked nonlinearities in the different categories considered separately. Plots (not shown here) of the numbers of publications as a function of laboratory size for each category, and corresponding plots of residuals from a least-squares line, suggest no systematic deviation from

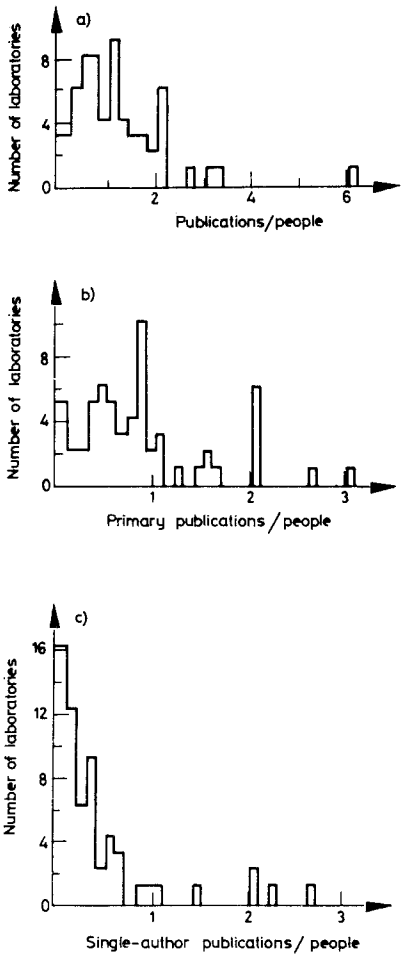

Fig. 2. The number of laboratories with each given ratio of (A) all publications to people; (B) primary publications to people; (C) single-author publications to people

linearity in any category. A few laboratories may have markedly more or markedly fewer publications than laboratories of comparable size in the same category. Fig. 2A is the frequency histogram of publication rate per capita, defined as number of publications/number of people, for all laboratories.

. *Per capita publication rate by category.* Laboratories in biochemistry and cell biology and in medicine and physiology have fewer additional publications per

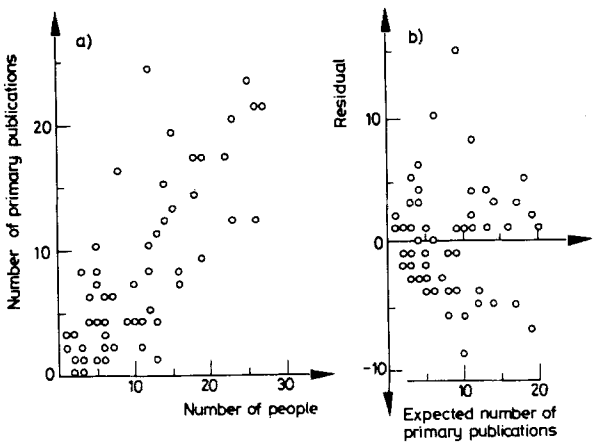

Fig. 3. For all laboratories, (A) the number of primary publications as a function of the number of individuals, and (B) the residual as a function of the estimated number of primary publications

capita than laboratories in chemistry, physics and mathematics, which in turn have fewer additional publications per capita than those in the behavioral sciences (Table 3). However, no statistically significant evidence for a difference among categories is found from a one-way analysis of variance, using either publication rate per capita or ln (5 + publication rate per capita) in order to reduce the non-normality of the data. Frequency histograms by category (not shown) also reveal no striking differences.

Primary publications and laboratory size. A scatter diagram of the number of primary publications as a function of laboratory size for all 60 laboratories (Fig. 3A) offers no suggestion of a non-linear relationship, nor does a plot of residuals (Fig. 3B). According to the fitted least squares line (method 1), primary publications = 0.596 + 0.709 X people. This line suggests that an additional 10 people produce seven additional primary publications per year per laboratory. Fig. 2B is a frequency histogram of the ratio, number of primary publications/number of people, for all laboratories.

The numbers of primary publications as a function of laboratory size plotted for each category (not shown) offer no suggestion of nonlinearity except possibly for the laboratories in medicine and physiology (category 3), where an apparent convexity may arise from fluctuation alone.

The fraction of all publications which are primary publications (= number of primary publications/number of all publications) for each laboratory, plotted as

Table 3
The increase in number of all publications and
in number of primary publications associated with an increase
by one person in laboratory size, for all laboratories and by category

Category	All publications measure of increase*			Primary publications measure of increase*		
	(1)	(2)	(3)	(1)	(2)	(3)
All	0.926	1.021	1.165	0.709	0.767	0.853
1	1.308	1.485	1.544	0.925	0.874	0.856
2	0.830	0.910	0.934	0.667	0.709	0.727
3	0.791	0.823	1.226	0.753	0.638	0.802
4	1.139+	1.275	1.391	1.179+	1.116	1.194

*(1) Slope of unconstrained least-squares line.

(2) Slope of least-squares line through origin, assuming variance of publications proportional to laboratory size.

(3) Slope of least-squares line through origin, assuming standard deviation of publications proportional to laboratory size.

+The slightly higher slope for primary publications arises because the estimated Y-intercept for all publications is slightly positive while for primary publications it is slightly negative.

function of laboratory size, reveals no systematic trend (not shown). We define the fraction of primary publications to be 1 for laboratories (of which there are only two) with 0 publications. The least-squared line (method 1) is: fraction of primary publications = $0.717 + 0.003 \times$ people. As the slope does not differ significantly from 0, the coefficients suggest that, regardless of size, approximately 7 out of 10 publications of a laboratory are primary publications.

Per capita primary publication rate by category. Laboratories in biochemistry and cell biology and in medicine and physiology have fewer additional primary publications per capita than laboratories in the behavioral sciences, which in turn have fewer additional primary publications per capita than those in chemistry, physics and mathematics (Table 3). Thus the marginal increase per capita for all publications is highest for the behavioral science laboratories but for primary publications is highest for chemistry, physics, and mathematics. The difference is due to a higher proportion of publications which are not primary (as defined here) in the behavioral sciences. However, a one-way analysis of variance reveals no significant difference in means among categories, using primary publication rate per capita or ln (5 + the primary publication rate per capita).

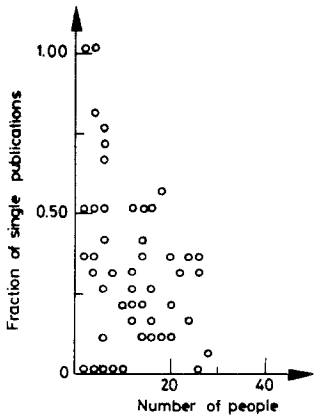

Fig. 4. The fraction of single-author publications (= number of single author publications/number of all publications) as a function of laboratory size, for all laboratories

Single-author publications. The fraction of single-author publications (= number of single-author publications/number of all publications) for each laboratory is plotted as a function of laboratory size in Fig. 4. For small laboratories, the fraction ranges from 0 to 1 (defining $0/0 = 1$ for laboratories with no publications). The largest fraction of single-author publications among laboratories of a given size systematically decreases as laboratory size increases. For laboratories with 20 or more people, not more than one-third of all publications have single authors.

Frequency histograms (not shown) of the number of single-author publications per capita for all laboratories and by category suggest the possibility that the laboratories in chemistry, physics, and mathematics (category 4) have more single-author publications per capita than the laboratories in biochemistry and cell biology (category 2). This possible difference is consistent with the evidence of Fig. 4 because the laboratories in category 4 all have 12 or fewer people, while those in category 2 include the largest laboratories.

Discussion

One major result of this study is that, at the Rockefeller University in 1977–1978, the number of all publications and the number of primary publications are directly proportional to the number of academic members of a laboratory. This proportionality holds for all laboratories considered together and, with one possible

exception, for laboratories separately in the behavioral sciences; biochemistry and cell biology; medicine and physiology; and chemistry, physics, mathematics and related sciences. For laboratories in medicine and physiology, there is a slight suggestion that the number of *primary* publications may be a convex function of laboratory size. In view of the overall pattern of simple proportionality, this suggestion should not be accepted without independent confirmation.

A priori, one could imagine that the bigger a research group, the better. Shared equipment, ideas, specialized books and journals, and common motivation could contribute to a publication rate per capita that increases with the size of the group. In this case, the number of publications as a function of group size would be increasing and upwardly convex.

On the other hand, one might suppose that the more time one spends talking with colleagues, the less time one spends thinking, discovering, and writing. An increase in the size of a group could mean an increase in the per capita administrative overhead necessary for the function of the group, since the number of possible pairwise relations increases as the square of group size. A plot of the number of publications as a function of laboratory size would, on these assumptions, be concave.

It is also plausible to suppose that economies of scale could accompany increases in size up to a certain point, after which diseconomies would dominate. If this were true, a plot of publications as a function of the number of individuals in a research group would be increasing and convex up to some size, at which marginal diseconomies began to dominate marginal economies, and would be concave thereafter, as in a logistic curve.

A fourth possibility is that there is a threshold size for a research group at which the need for explicit administrative organization and division of labor is first recognized. As a research group increases from a single individual up to this threshold size, the number of publications per year increases but not as fast as the number of people. Beyond this threshold an explicite division of labor and other synergistic effects produce an increasing per capita rate.

Finally, one might suppose that the number of publications of a group would be a linear function of group size. Such a relationship could arise if either the number of publications per year were an autonomous characteristic of individuals, independent of the size of group in which they worked, or if the economies and diseconomies of increasing group size were so nicely balanced that they cancelled each other out.

To explain the observed proportionality between the number of publications and the number of people in a laboratory, suppose that associated with each person is a random variable which gives the number of papers that person is capable

of producing in a year. If the person is one of n authors on the by-line of a single paper, this random variable measures that publication as 1/n of a paper and is thus identical to the "fractional productivity" of *Price* and *Beaver.*[6] Suppose also that this random variable is identically and independently distributed for every person in any category of laboratory. This model is neutrally consistent with the apparent association between the fraction of single-author publications and laboratory size (Fig. 4) because the model does not specify what fraction of a person's papers are to be singly authored. For the same reason, the model is neutrally consistent with the variation among fields of science in the average number of authors per published article.[7]

In addition to providing a (perhaps oversimplified) explanation of the proportionality between laboratory size and expected number of publications, the main use of the model is to suggest which of the three methods used in Table 3 may be most appropriate. Since the variance of a sum of independent random variables is the sum of the variance of each of them, this model implies that the variance in number of publications should be proportional to laboratory size, as assumed in method (2) in Table 3.

For every 10 members, method (2) suggests, there are 8 more publications per year in a laboratory of medicine or physiology, 9 more publications in biochemistry and cell biology, 13 more in chemistry, physics, or mathematics, and 15 more in the behavioral sciences. If one counts primary publications only, there are 6, 7, 11 and 9 more publications, respectively. Lest too much significance be attached to the apparent difference between categories in the proportion of all publications that are primary publications, we point out that our operational definition of a primary publication may not be uniformly appropriate for all categories of laboratories. In the behavioral sciences, for example, original research may be published in edited volumes relatively more often than in some other categories.

The lucid and presicent lectures of *Price*[11] mention not a single study of the effect of the size of a scientist's immediate working group on the productivity, however measured, of that group. The only prior observation we know of the relation between the size of a group of collaborating scientists and their aggregate rate of publication is based on groups in which authors are linked to one another through joint authorship of papers. In a study by *Price* and *Beaver,*[6] laboratory groups are not defined a priori; individuals who do not share authorship in at least one paper in a set of reports are excluded altogether. Two quantitative results of *Price* and *Beaver* may be compared to those obtained here. First, except possibly for the largest group of 77 authors, the ratio of the number of papers to the number of authors shows no clear increasing or decreasing trend as the number of authors in a group ranges from 1 to 58. Thus, as at Rockefeller University, the

number of papers per capita appears to be roughly independent of the size of the collaborating group. Second, the aggregate number of papers (533) divided by the aggregate number of authors (555) gives a quotient of 0.96 papers per author. This figure is not far from the ratio 631/618 = 1.02 observed at Rockefeller University. However, the papers studied by *Price* and *Beaver* are unpublished documents written over a five-year period, while the Rockefeller publications date from a single year.

The studies by *Pelz* and *Andrews*[8] do not relate the size of research groups to objective measures of group performance such as publications. However see Ref.[17]

The great importance of scientific publications as products and indicators of scientific work and the serious weaknesses of numbers of scientific publications as a measure of the quality, significance and impact of research are too well known to belabor here.[9,10] We use numbers of publications and number of primary publications to characterize aspects of the scientific output of laboratories because these numbers are objective, accessible, and easily understood.

Previous quantitative studies of scientific productivity using number of publications focus either on the distribution of number of papers per author[11,12] or on the evolution of the magnitude of highly aggregated scientific and technical literatures.[11,13]

To dramatize the difference between this study and studies of the number of publications per author, consider three hypothetical laboratories A, B, and C each with four people. Suppose, in laboratory A, one paper is published which is jointly authored by all four members of the laboratory. Suppose, in laboratory B, four papers are published, all by one member of the laboratory; the other three members publish nothing. Finally, suppose, in laboratory C, four papers are published and each paper is co-authored by all four members of the laboratory. Because the unit of analysis of this study is the laboratory, the per capita publication rate of laboratory A is 1/4 while laboratories B and C both have per capita publication rates of 4/4 = 1. If the unit of analysis had been the individual, and if we counted each paper on which the individual's name appeared as a publication, whether singly or jointly authored, then the average number of publications per individual in laboratory A would be $1 = (1 + 1 + 1 + 1)/4$, in laboratory B would be $1 = (4 + 0 + 0 + 0)/4$, and in laboratory C would be $4 = (4 + 4 + 4 + 4)/4$. Taking the laboratory as the unit of analysis gives laboratories B and C the same per capita publication rate. Taking the individual as the unit of analysis would give laboratories A and B the same average number of publications per individual.

Price[11] observes that an increase in the relative frequency of multi-authored papers is "one of the most violent transitions that can be measured in recent trends of scientific manpower and literature." Fortunately, his extrapolation that "at the present rate, by 1980 the single-author paper will be extinct" seems premature.

The example just given suggests that taking the laboratory as the unit of analysis for measuring productivity, by whatever index, may be one way of discounting appropriately an evident increase in the proportion of papers that individuals co-author.

Certain caveats accompany these results. The proportionality relation needs to be tested in other institutions. The required information is available, but not tabulated, for the National Institutes of Health[14] and possibly other research institutes. It is possible that proportionality holds only over a limited range of variation in laboratory size (here 1 to 27). A possible difficulty in interpreting data from laboratories significantly larger than those at Rockefeller is that a group listed as an administrative unit may not be an operating unit.

The relation of laboratory size to laboratory productivity needs to be examined using other measures of size and productivity. For example, another measure of the resources of a laboratory is its budget. Since the head of a laboratory is typically more highly paid than the postdoctoral fellows and graduate students who form most of the membership of larger laboratories, one might except the number of all or primary publications to be a convex function of the annual salary budget. On the other hand, larger laboratories require instrumentation and physical facilities which are often more expensive to provide and maintain than the equipment of smaller laboratories. One might expect the number of all or primary publications to be a concave function of total laboratory budget.

Many other indicators of scientific productivity have been proposed.[10] Some are based on citation frequencies, judgements by panels of peers, frequency of invited talks, and ability to attract extramural support. Citation frequencies have been avoided here because the numbers of citations to work in different fields may reflect primarily the total numbers of scientists in the different fields rather than quality.

For the management and administration of research institutions, information about productivity in relation to laboratory size could be useful. Clear evidence of a sharp advantage of either large or small laboratory size might stir administrative efforts to influence laboratory size in a direction favorable to scientific productivity. The proportionalities found here suggest that if the number of publications or the number of primary publications is accepted as a measure of productivity and the number of people is accepted as a measure of invested resources, there is no gain in productivity to be sought by favoring the investment of resources differentially according to laboratory size per se.

Differences in the sizes of laboratories may reflect differences in the style of work of different fields and of individual laboratory heads, how long a group has had to grow, and investment by funding sources within and outside the institution.

From these and other such factors it is difficult to derive a quantitative prediction of the distribution of laboratory size.

Caraco and *Wolf*[15] argue that the distribution of social group sizes is shaped primarily by ecological constraints. The mean sizes of prides of lions hunting various prey fall within the ranges of pride size which can meet lions' daily energetic requirements. This approach does not predict the form of the distribution of pride size but locates the center of the distribution. If each research laboratory is likened to a lion pride and each publication to a game kill (a unit of energy to be shared equally among the members of the group), then this ecological approach specifies no finite range of laboratory size because, at Rockefeller University, the publication rate per capita is independent of size. The use of other measures of the productivity per capita might make this ecological approach more helpful in accounting for laboratory size.

A second major finding of this study is that the observed frequency distribution of laboratory size is statistically indistinguishable from a 0-truncated negative binomial distribution. This theoretical distribution arises in a variety of biological data and can be derived in many ways. In particular, the negative binomial distribution is the equilibrium frequency distribution of size predicted by stochastic models of systems of freely-forming primate social groups.[4,16]

Perhaps the most appropriate of these models is model II of *Cohen.*[16] This model considers a collection of social groups (in this application, the laboratories in a research institution). Individuals may enter a group (or laboratory) from outside the institution, may leave a group to go outside the institution, or may migrate from one group to another in the institution. Arrival to a group is assumed to be described by two parameters. A parameter a describes the probability, per unit time per individual outside the institution, of attraction to a given group, regardless of the size of the group. This parameter summarizes the attractiveness of belonging to a group (at that institution) per se. A parameter b describes the attractiveness of a group per individual in the group. This attractiveness b of individuals is assumed to be the same for all individuals in the institution. The overall attractiveness of a group of size n to an individual outside the institution is the attractiveness of group membership per se plus the attraction of the n individuals in the group: $a + bn$. The probability of leaving a group to go outside the institution, per unit time per individual in the group, is described by a parameter d. Thus for a group of size n the probability per unit time of a departure to outside the institution is dn. For a migration from a group of size n to a group of size m within the institution, the probability per unit time is supposed to be $gnd(a + mb)$, where the constant g describes the intensity of intra-institutional migration.

Although there are four parameters a, b, d, g in this dynamic model, the equilibrium distribution of group size depends only on two ratios of parameters a/d and b/d which are related to the two parameters p and r of the truncated negative binomial distribution (see Table 2) by $a/d = r(1-p)$ and $b/d = 1-p$. (The parameter g does not affect the equilibrium distribution.) Fitting a negative binomial distribution to the observed frequency distribution of laboratory size yields in this instance the estimates $a/d = 1.804$, $b/d = 0.820$. The ratio of attractiveness of group membership per se to the departure rate is more than twice the ratio of the attractiveness of an individual in the group to the departure rate.

Finding that the frequency distribution of laboratory size is described by the negative binomial distribution does not confirm the dynamic assumptions of the model just described or of any other dynamic model. For such a confirmation, the time course of laboratory sizes would have to be studied.

In conclusion, the unit of analysis in this study is the laboratory, not the individual scientist. The size of a laboratory is measured directly by a de facto census of its academic members; size is not inferred from the authorship of papers. Because there is every incentive for members of a laboratory to report their publications, it is reasonable to suppose that all publications are enumerated; publications are not sampled by an abstracting or citation service. The directness of these measures of the numbers of scientists and numbers of publications probably contributes to the simplicity of the findings which emerge. These findings are that the number of publications of a laboratory in one year is proportional to the number of scientists in the laboratory during that year, and that the frequency distribution of laboratory size is statistically indistinguishable from the 0-truncated negative binomial distribution predicted by stochastic models for systems of social groups.

*

I thank Anne *Whittaker* for technical and editorial assistance at every stage of this work, which was partially supported by U.S. National Science Foundation grant DEB 74-13276. For helpful comments on a previous draft I am grateful to Einar *Gall*, Carl *Kaysen*, Rodney W. *Nichols*, Derek de *Solla Price*, George N. *Reeke*, Daniel *Simberloff*, and Burton *Singer*.

References

1. Rockefeller University. *1977–1978 Catalogue.* New York, Rockefeller University, 1977.
2. G. W. SNEDECOR, W. G. COCHRAN, *Statistical Methods.* 6th ed. Ames, Iowa State University Press, 1967, p. 168–169.
3. Rockefeller University. *Annual Report July 1977–June 1978.* New York, Rockefeller University, 1978.

4. J. E. COHEN, *Casual Groups of Monkeys and Men: Stochastic Models of Elemental Social Systems.* Cambridge, MA, Harvard University Press, 1971.
5. J. W. TUKEY, *Exploratory Data Analysis.* Reading, MA, Addison-Wesley, 1977, Chapter 5.
6. D. de SOLLA PRICE, D. deB. BEAVER, Collaboration in an invisible college. *American Psychologist,* 21 (1966) 1011–1018.
7. D. deB. BEAVER, R. ROSEN, Studies in scientific collaboration III. Professionalization and the natural history of modern scientific co-authorship. *Scientometrics,* 1 (1979) 231.
8. D. C. PELZ, F. M. ANDREWS, *Scientists in Organizations.* Rev. ed. Ann Arbor, Michigan, Institute for Social Research, The University of Michigan, 1976.
9. J. M. ZIMAN, *Public Knowledge: An Essay Concerning the Social Dimension of Science,* Cambridge, University Press, 1968.
10. Y. ELKANA, J. LEDERBERG, R. K. MERTON, A. THACKRAY, H. ZUCKERMAN, *Toward a Metric of Science: The Advent of Science Indicators,* New York, J. Wiley, 1978.
11. D. de SOLLA PRICE, *Little Science, Big Science,* New York, Columbia University Press, 1963.
12. D. de SOLLA PRICE, A general theory of bibliometric and other cumulative advantage processes. *Journal of the American Society for Information Science,* 27 (1976) 292–306.
13. D. W. KING, *Statistical Indicators of Scientific and Technical Communication, 1960–1980,* Vol. 1. A Summary Report for the National Science Foundation Division of Science Information. Washington, D. C.: U.S. Government Printing Office, 1976.
14. National Institutes of Health, *Scientific Directory 1978; Annual Bibliography 1977,* DHEW Publication (NIH) 78–4. Bethesda, MD; National Institutes of Health, Division of Scientific Reports, 1978.
15. T. CARACO, L. L. WOLF, Ecological determinants of group sizes of foraging lions. *American Naturalist,* 109 (1975) 343–352.
16. J. E. COHEN, Markov population processes as models of primate social and population dynamics. *Theoretical Population Biology,* 3 (1972) 119–134.
17. Pertinent results are reported and cited by Rikard Stankiewicz, "The size and age of Swedish academic research groups and their scientific performance," Chapter 8 in *Scientific Productivity: The Effectiveness of Research Groups in Six Countries,* Frank M. ANDREWS (Ed.), Cambridge University Press and Unesco, 1979.