Scientometrics, Vol. 4. No. 5. (1982) 361–378

DETERMINANTS OF EXPERT JUDGEMENT OF RESEARCH PERFORMANCE

M. E. D. KOENIG

School of Library Service, Columbia University, New York, NY 10027, (USA)

(Received February 24, 1982)

The relationship between indicators of, and expert judgement of, research performance were compared in the context of mission oriented pharmaceutical research. Expert judgment is very highly correlated with measures of publication activity, much more so than with very plausible measures of research output and research quality. Furthermore, expert judgement appears to be an additive function of publication size (another name for which might be visibility) and publication quality, with the principal component being size/visibility. These results are very similar to those found by *Anderson, Narin,* and *McAllister* in the context of academic research,[1] but these findings emerge from a context which allows other variables to compete in predicting expert judgement, and are therefore to that degree more robuts. In addition this study finds a clear pattern of subject specificity, which implies that visibility is a function of the judge's subject field.

Introduction

In undertaking a scientometric study of pharmaceutical research, the author was presented with the opportunity of examining the relationship between publication indicators and expert judgement of research performance. The results of this analysis amplify and extend previous work on the topic.

The principal previous work is that of *Anderson, Narin,* and *McAllister* which indicated strongly that expert judgements of the quality of research, specifically the quality of graduate faculties of U.S. academic institutions, are constituted of "two additive components, bibliometric* size and bibliometric quality".[1] In this study of misson oriented pharmaceutical company research, the findings arrived at were strikingly parallel and corroborative. These results reinforce and extend the previous findings in two directions. First, the results obtained in this study emerged from a

*Although *Anderson* et al[1] use "bibliometric size" and "bibliometric quality" as a generic term we fell that the word publication instead of bibliometric is more adequate and are using it throughout this paper.

statistically far more complex environment and therefore they argue more persuasively for the scientometric nature of the dependence. Second, the results exhibit a very plausible relationship to the subject expertise of the judges – in other words, the publication* size and publication quality components appear to be a function of the subject specialization of the expert making the judgement.

To review the previous work, *Anderson, Narin* and *McAllister* compared *Roose-Andersen* (expert judgement) data with publication data from the same subject fields. The number of universities, for each subject field, varied from 55 to 88. The *Roose-Andersen*[2] data derive from a study conducted in 1969 as a follow up to an earlier study conducted by *Cartter*[3] for the American Council on Education. The study was a survey analysis of informed opinions of over 6,000 respondents asked to note, for their own field, the quality of the graduate faculty and the effectiveness of the doctoral program of some 125 U.S. academic institutions. Against these scores, *Anderson* et al, analyzed three publication variables: Total number of papers, a publication size measure; Influence per Paper,* a size independent measure of citation influence, and Total Influence, defined as total papers x influence per paper.

The conclusion of *Anderson* et al is that partial correlations and regression analyses "indicate that Roose-Andersen scores have two additive components: publication size and publication quality," with publication size being the primary component.

Context

In the study reported here, the research performance of nineteen major pharmaceutical companies** was examined. Pharmaceutical research is a particularly attractive area in which to study the relationship of various indicators of research performance. The principal attraction is that pharmaceutical research has indicators of research output that are quantifiable, reasonably commensurate, and publicly accessible. Those indicators are the phamaceutical agents themselves, discrete and reasonably numerous. In addition the complex NDA (New Drug Application) approval process ensures that they are examined and evaluated in a very public arena, one that is

*Influence Per Paper is analogous to *Garfield's* Impact Factor, the number of citations a journal receives normalized by the number of papers it publishes, or its opportunity to be cited. The Influence Per Paper extends the impact factor, by calculating on a field by field basis, and by iteratively assigning more weight to citations from the journals which are themselves more heavily cited.

**The basis for inclusion was that the bulk of the companies research be done in North America (to limit commensurability problems in R&D budget data) and that the company have developed and marketed in the years 1965–1976, at least one drug judged an important therapeutic gain by the Food and Drug Administration.

subject to intense Monday-morning quarterbacking. The approval of an NDA is probably as unrelated to scientific prestige or visibility, probably as directly a function of efficacy and safety as one is likely to find in this imperfect world. In addition, the sheer magnitude of the process ensures that no pharmaceutical company undertakes an NDA for a trivial reason. The NDA process sets a minimum or threshold size on the extent of the R&D process needed to support an NDA. This threshold generates its own controversy,* but for the purpose of this study it very usefully enforces a certain commensurability of size. Finally, the NDA process results in public accessibility and provides evaluative data, to be described below, generated by the FDA itself.

A second major advantage is the fact that the pharmaceutical industry is relatively unconcentrated. That is, there are relatively a large number of pharmaceutical companies, so that one can deal with a number of cases sufficient to have a reasonable expectation of arriving at results with some statistical validity. Furthermore, the industry is relatively stable over time in terms of entrances, departures, mergers, etc., so that it is a relatively tractable area to study over time.

Although uniqueness is always a dangerous claim, the combination of the features above may very well make pharmaceutical research a unique arena in which to study the research process.

The advantage for this study is that pharmaceutical research provides a rich menu of research indicators in which to examine the relationship between expert judgements and bibliographic data. Perhaps most important, it allows the introduction of other indicators of research performance, other independent variables in analyzing expert judgement.

Data

For the nineteen pharmaceutical companies in this study, four major data aspects or categories were examined. Those four categories are:

1. Expert judgement data as to the research performance of the companies.
2. Publication variables: articles produced by those companies, and citation data to those articles.

*The controversy is whether the rigidity and the complete Go/NoGo approach of the FDA, coupled with the size of the effort involved in obtaining an NDA, effectively prevents pharmaceutical companies from developing drugs for disease areas where there may be real need, but a low incidence and therefore a small market.

3. Research output in the form of new pharmaceutical agents approved by the Food and Drug Administration.
4. Organizational size in the form of R&D budgets for pharmaceutical companies.

The data is discussed in more detail below. The data for all nineteen companies is complete; there is no missing data.

Expert Judgements

The judgements concerning the research performance of the pharmaceutical companies were collected by surveying a panel of experts, the members of the NIH (National Institutes of Health) export advisory panel on pharmacology.

The committee members, 28 in number, were surveyed by mail, and asked to rate the phamaceutical companies in regard to four specific research criterion:

1. "creativity and innovativeness in their pharmaceutical research" – Creativity
2. "overall contribution to medical well-being" – Contribution
3. "commercial effectiveness in capitalizing upon pharmaceutical research" – Commercialization
4. success "in pursuing basic biomedical research" – Basic Research

Fifteen responses were received, for a return rate of 54%. Since the respondents were guaranteeed anonymity, non-respondents were not distinguishable, but an attempt was made to phone each questionnaire recipient to remind that person to please complete the form. This technique is probably at least partially responsible for the reasonable response rate.

The recipients were asked to circle the names of approximately five of those companies most characterized by the criterion, and to *underline* the names of approximately five of these companies least characterized by it. This technique was felt to provide reasonable discrimination, while at the same time not being unduly onerous for the respondent, and thereby discouraging response.

Although the number of responses is low in absolute terms, the consistency of those responses gave confidence that very little more information would have been gained by a larger sample size or by a higher response rate. There was extremely high agreement at both ends of the scale; contradiction, a company receiving both positive and negative votes on a criterion, occurred virtually only in the middle ranges. That is, a company receiving a large number of positive votes would typically receive some neutral votes (neither a positive nor a negative vote), but no negative votes, and conversely a company receiving a large number of negative votes, typically received some neutral votes and no positive votes.

The raw scores were compiled by simply assigning a score of $+1$ for each positive characterization and a score of -1 for each negative characterization. This

Table 1
Correlations of expert opinion variables

	Expop	Creativity	Contribution	Basic research
Expop				
Creativity	0.99			
Contribution	0.98	0.95		
Basic Research	0.96	0.93	0.90	
Commercialization	0.69	0.66	0.72	0.62

(All correlations are significant at the 1% level)

resulted in a range of score from +15 to −9. The scores were then scaled up by adding 10 points to each, yielding a possible range from 25 to 1.

The responses to the Creativity, Contribution and Basic Research questions were very highly correlated to each other, see Table 1. The answers to the Commercialization question fell into a different pattern. The variable *Expop (Expert Opinion)* was created for analytical purposes simply by summing the scores of Creativity, Contribution, and Basic Research, to avoid the repetition in subsequent analyses of three almost identical measures.

The scores for the nineteen companies are presented in Table 2. (pharmco = pharmaceutical company)

Publication Variables

The basic publication data consisted of the articles published by the nineteen major pharmaceutical companies, and citations to those articles. Publication data for the 19 companies were gathered for the years 1970 to 1974. The data source was the Corporate Index to the *Science Citation Index*[4] of the Institute for Scientific Information. *The Science Citation Index* processes as source material all articles appearing in some 2 500 core scientific journals. Each years' articles are sorted and arranged by organization in a corporate index.

The citation data were compiled directly from *Science Citation Index*. Each pharmaceutical company article was searched in the annual Citation Index of the year three years subsequent to the appearance of the source article. That is, a 1970 article would be searched in the 1973 Citation Index, a 1974 article would be searched in the 1977 Citation Index. The three year time lag was chosen to maximize the opportunity for an article to be cited, as previous research has revealed that the typical article reaches its peak citation rate after approximately three years.

Table 2
Pharmco expert opinion scores

Pharmco	Creativity	Contribution	Commercial-ization	Basic research	*Expop*
Abbott	5	5	10	6	16
Ayerst	5	4	15	8	17
Bristol	6	9	7	10	20
J & J	7	8	16	7	22
Lederle	10	15	8	12	37
Lilly	20	19	15	19	58
Mead J.	6	4	7	10	20
Merck	24	25	24	24	73
Pfizer	7	7	10	9	23
Roche	22	23	23	23	68
Schering	7	6	12	7	20
Searle	7	5	4	3	15
Smith K.	18	18	20	13	49
Squibb	8	9	7	6	23
Sterling	1	5	7	8	14
Syntex	8	9	12	10	27
Upjohn	22	17	7	23	62
Warner L.	2	3	8	3	8
Wyeth	4	4	6	4	12

The articles were classed into four classes:
- biological
- clinical medicine
- chemistry
- biomedical research

The basis of the categorization was the subject field of the journal in which the article appeared. To categorize the journals, a classification scheme developed principally by Dr. Gabriel Pinski of Computer Horizons was used. The classification scheme is described and set forth in *Evaluative Bibliometrics.*[5]

In addition, the articles were classed into four citation classes: uncited articles; singleton articles, those articles cited only once in the third year; journeyman articles, those cited two to four times in their third year; and star articles, those cited five or more time in their third year. A cross tabulated dispaly of the sixteen categories that resulted is presented in Table 3 below. In addition to these sixteen variables, publication variables for each company were created as follows, cumulat-

Table 3
Cross-tabulated display of citation frequency class by subject area of article

Count & Column Percent	Subject area of article				
	Biology	Clinical Medicine	Chemical	Bio-Medical research	Row Total
Citation Frequency Class Uncited (0 Cites)	270	2325	1010	1145	4895
	66.0%	46.9%	53.1%	48.8%	49.9%
Singletons (1 Cite)	74	987	369	372	1821
	18.1%	19.9%	19.4%	15.9%	18.6%
Journeymen (2–4 Cites)	55	1118	374	444	2005
	13.4%	22.5%	19.7%	18.9%	20.5%
Stars (> 4 Cites)	10	532	150	386	1082
	2.4%	10.7%	7.9%	16.4%	11.0%
Column Total	409	4962	1903	2347	9803
	4.2%	50.6%	19.4%	23.9%	100.0%

(Note, row total column includes 1.8% miscellaneous articles.)

ive numbers of articles and citations in each of the four subject classes, total articles, total citations, and citation rate (total articles/total citations), for a total of 27 publication variables.

Research Output

The basic measure of pharmaceutical company research output is taken here as the number of NDAs (New Drug Applications) approved by the FDA (Food and Drug Administration) for each company during the years 1965 to 1976. While the approved NDA is by no means a perfect measure of research output, it does have a high degree of intuitive validity: As the phrase in the pharmaceutical industry has it "the NDA is the name of the game." From the point of view of the pharmaceutical company, its R & D effort is pointless if no approvable NDAs emerge. An approved NDA, that is a marketable drug, is the basic goal of pharmaceutical company research; all else is ancillary.

This measure however was refined in three ways. First, only those NDAs that were new chemical entities were considered. Medical devices, combination products, new uses of an existing compound and minor molecular modifications were ex-

cluded. The rationale for this distinction is that it reasonably reflects the distinction between those NDAs that represent original research and those which are generally derivative. The correspondence is not perfect, but this measure has the advantage that it is simple, objective, and replicable. Furthermore, it has been used in previous studies of pharmaceutical innovation and has general scholarly acceptance.[6,7] The definition of "new chemical entity" is that of Paul deHaen Inc., used in their various commercially available drug research information products, particularly the *New Product Parade.*[8] The deHaen definition is widely accepted and has been used in various studies of the pharmaceutical industry, including use by the the FDA itself.[7]

Second, those drugs on which the company enjoyed patent protection were weighted more heavily, on a ratio of 2 to 1, than those drugs without. The rationale for this was that, if a pharmaceutical company has done the development work on a compound, it will generally enjoy patent protection. If it has not done the developmental work, it will typically not enjoy patent protection. In those cases, the patent may have been assigned to the company from whom the patent was licensed, or the discovery may have originated in the open literature. This study used the patent selection made by the editors of the Merck Index[8] in the compilation of "monographs" on the drugs in question. Each monograph is a capsule history of the drug, rather than a therapeutic guide, highlighting its chemical, pharmacological, and medicinal properties, and contains references to key papers and key patents. If in the Merck Index there is a patent assigned for a particular drug to the pharmco which has been granted the NDA, that drug will be classed as having patent protection, otherwise not.

Third, the drugs were weighted by medical importance. This weighting is based on an internal FDA (Food and Drug Administration) study which classified drugs as to therapeutic potential. Those drugs regarded by the FDA as being particularly important were classified as "Important Therapeutic Gains". Such drugs, ITGs, were weighted in this study in a ratio of $2^1/_2$ to 1 to non ITGs, which ratio was derived simply from the inverse of the relative proportion of ITGs to non-ITGs.

These refinements and weightings resulted in a component drug output score, and four constituent components. Those components were:

Important Therapeutic Gain drugs with company patent positions (ITG–P)
Important Therapeutic Gain drugs without company patent positions (ITG–WP)
Modest Therapeutic Gain drugs with company patent positions (MTG–P)
Modest Therapeutic Gain drugs without company patent position (MTG–WP)

The composite drug output can be expressed as:

$$\text{Score} = (5 \times \text{ITG–P}) + (2.5 \times \text{ITG–WP}) + (2 \times \text{MTG–P}) + (1 \times \text{MTG–WP}).$$

In Table 4 below, the drug output data is presented.

Table 4
Pharmaceutical companies ranked by drug output, ITGs (Important Therapeutic Gains), and score

Rank	Drugs/Pharmco	ITG/Pharmco	Score/Pharmco	Rank
1	12 Pfizer	5. Pfizer	63.00 Pfizer	1
2	12 Roche	5. Merck	53.50 Upjohn	2
3	10 Upjohn	4. Roche	48.00 Roche	3
4	8 Warner L.	4. Upjohn	47.00 Merck	4
5	8 Merck	3. Ayerst	33.50 Warner L.	5
6	8 Lilly	3. Abbott	28.00 Sterling	6
7	7 Bristol	2. Warner L.	26.50 Bristol	7
8	6 Squibb	2. J & J	24.50 J & J	8
9	6 Schering	2. Bristol	22.50 Squibb	9
10	5 Öederöe	2. Schering	21.50 Lilly	10
11	5 J & J	2. Sterling	19.50 Schering	11
12	5 Abbott	2. Squibb	17.00 Abbott	12
13	4 Ayerst	1. Lilly	12.50 Lederle	13
14	4 Smith K.	1. Syntex	10.50 Syntex	14
15	4 Sterling	1. Smith K.	8.50 Ayerst	15
16	3 Syntex	1. Lederle	6.00 Mead J.	16
17	3 Mead J.	0. Searle	5.50 Smith K.	17
18	2 Wyeth	0. Mead J.	4.00 Searle	18
19	1 Searle	0. Wyeth	2.00 Wyeth	19

Research Input

Financial data on the size of pharmaceutical companies' research efforts were collected for the years 1965 to 1978, to serve as a measure of the size of the input to a pharmco R&D effort. Dollars of research budget have the advantage of being relatively commensurable and relatively obtainable. While other measures of the size of research effort are possible, the number of people employed for example, they present difficulties. Data on the number of people employed is not easily obtainable, nor is it consistent – what one company may describe as R&D personnel, for example programmers and analysts supporting the R&D efforts, may be assigned quite differently by another company.[10] Furthermore, the ratio of people to resources deployed is not consistent, and thus people may not be an ideal indicator of resources deployed. One company may have invested far more heavily in laboratory automation than another, a third may subcontract much of its biological screening.

Financial data present some difficulties however. One is the comparability of different currencies over time, another is the commensurability of actual purchasing power in different countries. How does one deal with changing exchange rates and differing salary ranges in various countries? This class of problems was one of the reasons for limiting the analysis to pharmcos doing the bulk of their R&D in North America; another reason was the difficulty of obtaining financial data for pharmcos in Europe, where financial reporting requirements are generally less rigorous than in the U.S., and where closed ownership is more frequent.

The basic source for pharmco R&D budget data are the annual reports and the 10K reports filed with the Securities and Exchange Commission (for those companies publicly traded in the U.S.). Numerous complications remain however; for example, R&D data have been required in 10K reports only since 1973, and even now it is not yet required that companies break down their R&D by product line. Furthermore, if a company, such as American Home Products, has functionally and organizationally separate subsidiaries, such as *Wyeth* and *Ayerst*, financial data need not be broken down by subsidiary.

The data shown below in Table 5 is based on figures from annual reports and 10Ks where available, but they have been modified and refined as a results of numerous phone calls and conversations with the author's former colleagues in the pharmaceutical industry, many of whom were gracious enough to volunteer information whose release was clearly in violation of company policy. Needless to say, these sources cannot be acknowledged. These sources were used to disambiguate situations such as the above example of *Ayerst* and *Wyeth*, and to extract the not relevant R&D costs, such as consumer products R&D (for companies such as Bristol Myers, those expenses are a significant component of the reported R&D costs) from the reported figures. The author is quite comfortable with the data with one exception, that being that the data for Hoffmann La Roche seems surprisingly low. The numbers were reported with apparent sincerity by a Hoffman La Roche executive whose job position included not only a familiarity with the numbers, but a considerable degree of responsibility for them. Because of the suspicion that these numbers are low, the major analyses were repeated using an average budget for Hoffman La Roche twice that reported. This fortunately did not show any noticeable effect on any of the analyses.

The inclusion of research input data also allowed the creation of an output/input variable, or a "quality" measure of the companies' performance. This variable was labelled "productivity", and was calculated as (Drug Output Score/Average Yearly R&D Budget) $\times$ 100. The productivity data is presented in Table 6 below.

Table 5
Pharmaceutical companies ranked by average research budget

Rank	Pharmco	Average yearly research budget (in millions)
1	Merck	87.0
2	Lilly	76.7
3	J & J	67.2
4	Warner Lamber	58.1
5	Upjohn	57.3
6	Roche*	56.4
7	Pfizer	52.3
8	Smith Kline	40.8
9	Abbott	37.2
10	Squibb	35.7
11	Searle	33.8
12	Schering	31.6
13	Lederle	27.5
14	Wyeth	26.0
15	Bristol	25.6
16	Sterling	23.6
17	Syntex	15.3
18	Mead Johnson	11.7
19	Ayerst	10.9

Table 6
Pharmaceutical companies ranked by productivity (Score/AvgBud)

Rank	Productivity	Pharmco
1	12,0	Pfizer
2	11.9	Sterling
3	10.5	Bristol
4	9.4	Upjohn
5	8.5	Roche
6	8.3	Ayerst
7	7.2	Syntex
8	6.4	Squibb
9	6.3	Schering
10	5.9	Warner L.
11	5.4	Merck
12	5.1	Mead J.
13	4.7	Lederle
14	4.6	Abbott
15	3.7	J & J
16	2.9	Lilly
17	1.5	Smith K.
18	1.18	Searle
19	0.8	Wyeth

*Note comments

Analysis

In examining the correlations between Expert Judgement and various independent variables, it is quickly apparent that the highest correlations are with publication variables, in particular biological articles. (Tables 7, 8)

The four expert judgements, Creativity, Contribution, Basic Research, and Commercialization, as well as the composite variable *Expop*, were submitted to an order unconstrained regression analysis, using SPSS.[11] The results of those equations, carried only as far as three variables, are presented, along with the coefficients (Beta weights), the Adjusted R Square and the partial F scores for each step. The variables have been normalized (that is, the dependent and independent variables have been standardized to have unit variance and a mean of zero), and consequently the coefficients used are Betas, and there is no constant. The intent of this normalization is to better show the relative contributions of the different variables. In ad-

Table 7
Correlation coefficients with *Epop* (Creativity + Contribution + Basis Research) and related variables

Bibliometric Variables		Other Variables	
All Articles	0.75**	Score (Drug Output)	0.48*
Star Articles	0.70**	Important Therapeutic Gains	0.44*
All Citations	0.69**	Productivity	0.00
Biological Articles	0.84**	R&D Budget	0.62**
Clinical Articles	0.72**		
Star Biological Articles	0.57**		
Star Clinical Articles	0.77**		

*Significant at the 5% level
**Significant at the 1% level.

Table 8
Correlation coefficients with Commercialization and relate variables

Bibliometric Variables		Other Variables	
All Articles	0.53**	Score (Drug Output)	0.27
Star Articles	0.56**	Important Therapeutic Gains	0.40*
All Citations	0.55**	Productivity	–0.06
Biological Articles	0.61**	R&D Budget	0.40**
Clinical Articles	0.42**		
Star Biological Articles	0.52**		
Star Clinical Articles	0.33**		

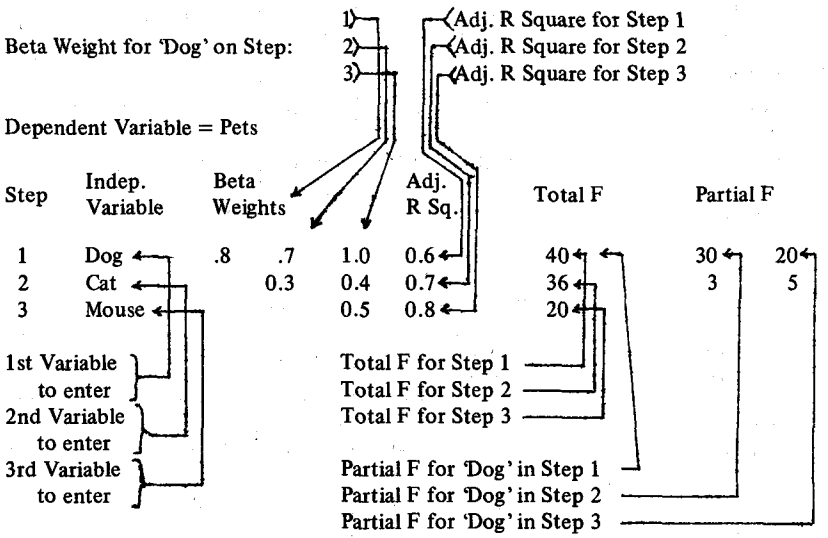
*
**See Table 7

dition, the first three independent variables awaiting entry into the equations, at steps 1 and 2 are presented, with their partial Betas and partial F scores.

The tableau format used is illustrated in Table 9. The correlation coefficients for *Expop,* Creativity, Contribution, Basic Research and Commercialization are shown in Tables 10–14.

What is immediately apparent about these equations is their consistency: In all five cases, the first independent variable is Biological Articles. In three of the five cases, Creativity, Contribution, and *Expop* (the sum of Creativity, Contribution, and Basic Research), the second independent variable, is Star (highly cited) Biological Articles. In a fourth case, Basic Research, the second independent variable is Journeymen Biological Articles.

Table 9
The format of variables



Step 1–Independent variable = dog
Step 2–Independent variables = dog + cat
Step 3–Independent variables = dog + cat + mouse

Table 10
Expop as dependent variable

Step	Indep. Variable	Beta Weights			Adj. R Sq.	Total F	Partial F	
1	Articles-Biological	0.84	0.73	1.04	0.69	41.0**	26.8**	21.7**
2	Stars-Biological		0.24	0.34	0.72	24.2**	2.9	5.4*
3	Journeymen-Clinical			-0.42	0.75	19.1**		3.0

Only in one case, that of Commercialization, does the second independent variable differ markedly; it is Important Therapeutic Gains without Patent Position, Even this result, apparently anomalous, is perhaps an exception that proves the rule.* Commercialization, is indeed the one question least directly related to the

*In this expression, 'prove' is of course the gunmakers prove.' as in proofmark; it is not the mathematician's or statistician's 'prove'.

Table 11
Creativity as dependent variable

Step	Indep. Variable	Beta Weights			Adj. R Sq.	Total F	Partial F	
1	Articles-Biological	0.84	0.70	1.16	0.70	42.4**	28.6**	12.6*
2	Stars-Biological		0.30	0.36	0.76	28.8**	5.1*	7.2*
3	Cites-Clinical			−0.52	0.77	21.5**		2.3

Table 12
Contribution as dependent variable

Step	Indep. Variable	Beta Weights			Adj. R Sq.	Total F	Partial F	
1	Articles-Biological	0.79	0.68	1.04	0.61	29.0**	17.9**	17.1**
2	Stars-Biological		0.25	0.36	0.64	17.0**	2.4	5.0*
3	Journeymen-Clinical			−0.49	0.68	13.9**		3.2

Table 13
Basic research as dependent variable

Step	Indep. Variable	Beta Weights			Adj. R Sq.	Total F	Partial F	
1	Articles-Biological	0.87	0.73	1.19	0.74	30.9**	26.4**	14.2**
2	Journeymen-Biological		0.24	0.29	0.77	27.6**	6.2*	4.3*
3	Cites-Clinical			−0.62	0.78	24.2**		2.4

Table 14
Commercialization as dependent variable

Step	Indep. Variable	Beta Weights			Adj. R Sq.	Total F	Partial F	
1	Articles-Biological	0.61	0.46	0.93	0.34	10.1**	6.3**	8.1**
2	ITG without . . .		0.43	0.37	0.48	9.1**	5.6*	4.5*
3	Journeymen-Clinical			−0.53	0.53	7.8**		2.9

quality of research itself, and it is the question to which the judges' response was most singular. It is not at all implausible that drugs without patent positions, presumably drugs discovered externally to the company should be associated with commercialization. The ability to successfully and competitively recognize externally developed compounds of significant and marketable therapeutic potential, would clearly correspond to successful commercial acumen. This relationship in context with the clear distinction between the response to commercialization versus the other three questions, gives a certain credibility to the expert judgement responses, and a certain confidence that the responses were a result of thoughtful deliberation.

Overall, there are two salient aspects that emerge. The most obvious, about which more below, is the consistency of appearance of biological articles. The second salient aspect is the pattern of the first independent variable being an aspect of publication size, number of articles, and the second being an aspect of publication quality, the star or highly cited articles. The responses more directly related to research quality, Creativity, Contribution, and the summary variable *Expop* follow this pattern precisely, and Basic Research, merely substitutes journeymen biological articles for star biological articles. These too, it can be argued, are not inappropriate as an indicator of bibliometric quality; a citation rate of 2–4 cites a year, even in the highly cited third year, puts a biomedical article in at least the top quartile by citation rate.[1]

There is a very strong resemblance between these results and the results obtained by *Anderson, Narin,* and *McAllister* in their bibliometric analysis of Roose-Andersen scores of graduate departments, in which they come to a very parallel conclusion that the *Roose-Andersen* (expert judgement) scores have two additive components: publication size (number of publications) and publication quality ("influence per paper" a citation count with citation weights based on the citation rate of the citing journal, determined in an iterative fashion).[1] One might similarly describe the regression results above as displaying two additive (both positive in sign) components, a publication size component (number of biological publications) and a bibliometric quality component (number of highly cited biological publications).

The results of the analysis of the study here however are much more robust than those of the previous study. The *Anderson, Narin, McAllister* study compared expert judgement only with publication data. There were no other competing independent variables. Here, however four component indicators and one composite indicator of research output in the form of therapeutic agents, an indicator of research organizational size in the form of R&D budget, an output/input indicator of research productivity or quality, plus a much more extensive menu of

some 27, rather than 3 publication measures, were also included, and a very comparable relationship was found.

Of additional interest is the consequence of the first observation above, the almost exclusive presence among the independent variables of the regression equations, of biological publication variables. This two step additive process then seems to be very subject specific, based almost entirely, for these judges, on biological articles, rather than on all articles. The second variable is of course chosen based not on its direct correlation with the subject variable, but based on its explanatory power in terms of the variation remaining "unexplained" by the first variable. Even so, the fact that Biological Stars is typically the second variable, and not other variables, such as Clinical Stars, which have higher first order correlations, is thought provoking. The fact that the size and the quality components are in the same subject field lends some face validity to the result. It clearly seems to imply that our panel of experts was particularly influenced by biological visibility. Althouth pharmacology, the area of expertise of the panel of judges, is typically thought of as clinical medicine, the basic science from which it derives is physiology, a biological science. In the classification scheme used, pharmacology journals are indeed classifie as clinical, and physiology journals are classified as biological.

Table 15
Expop as dependent variable

Step	Indep. Variable	Beta Weights			Adj. R Sq.	Total F	Partial F	
1	[AvgBud]	0.63	0.18	0.32	0.36	11.1**	1.2	3.3*
2	Articles-Biological		0.73	1.30	0.69	21.4**	19.6**	12.7**
3	Citations Clinical			–0.69	0.73	17.1**		3.0

Table 16
Commercialization as dependent variable

Step	Indep. Variable	Beta Weights			Adj. R Sq.	Total F	Partial F	
1	[AvgBud]	0.53	0.41	0.27	0.23	6.5*	5.0**	2.2
2	ITG without . . .		0.49	0.42	0.44	8.1*	7.3**	5.9**
3	Journeymen-Clinical			–0.34	0.51	7.2*		3.2

[ ] = variable constrained to appear in that position.

To examine for the impact of scale effects, the expert judgement variables were also submitted to a regression equation in which Average R&D Budget was constrained to enter first. Average R&D Budget is of course a reflection of organizational size. These regression equation tableaus for *Expop* and Commercialization are presented in Tables 15, 16.

The importance of biological articles, publication visibility as it were, clearly remains. The increase in adjusted R^2 as Biological Articles enters on step two is clearly apparent. The importance of 'Important Therapeutic Gain Drug without Patent Protection' to Commercialization also remains evident.

In examining the regressions above, the clear implications, both from the ordering and from the adjusted R squares, is that publication size is far more important than publication quality. Although publication quality is an additive component (and here the consistency of subject, as biological star articles enter on the second step well ahead of other variables with higher first order correlations, is persuasive) it indeed appears to be far less important than publication size or visibility. While the *Anderson, Narin,* and *McAllister* study is not completely comparable, resting its case primarily upon partial correlation rather than upon stepwise regression, their data indicates a similar ordering of importance. This study, however, indicates a more predominant rule for publication size than the previous study suggests.

Summary

In summary, the data demonstrates that expert judgement of research performance is very highly correlated with measures of publication activity, and that expert judgement appears to be an additive function of publication size (another name for which might be visibility) and publication quality, with the principal component being size/visibility. Even when organization size is controlled for, the importance of publication visibility remains paramount.

Furthermore, here as in the previous study by Anderson et al, the expert judgements are very predictable from publication data. What this study adds, is convincing evidence that this predictability is greater from publication variables than it is from other indicators of research performance. While the bullpen of research indicators with which the publication variables competed in this study is by no means exhaustive, it is quite extensive. Despite the competition as it were, the bibliometric variables emerged as the overriding factors in predicting expert judgement of research performance.

Additionally, there emerged a clear pattern of subject specificity. The publication variables selected by the regression equation were consistently of the same sub-

ject classification, a subject consistent with the academic and research interests of the panel of judges. This suggests very strongly that publication visibility, and therefore expert opinion of research performance, is a function of the judge's exposure to the literature, in which of course the literature of the judge's subject field predominates.

Further work on this topic suggests itself directly. With the use of a finer classification scheme such as the subject specialty clusters developed by *Small* of ISI,[12,13] and with the use of co-citation techniques both the pharmaceutical company output and the judges themselves could be assigned to a far more precise subject map. This would allow far finer analysis of the relationship between publication exposure and expert opinions of research performance.

References

1. R. C. ANDERSON, F. NARIN, P. McALLISTER, "Publication Ratings versus Peer Ratings of Universities", *Journal of the American Society for Information Science*, 29 No. 3 (1978) 91–103.
2. K. D. ROOSE, C. J. ANDERSON, *A Rating of Graduate Programs*, American Council on Education, Washington, D.C., 1970.
3. A. M. CARTTER, *An Assessment of Quality in Graduate Education*, American Council on Education. Washington, D.C.
4. *Science Citation Index*, Philadelphia, Pa., Institute for Scientific Information, 1961 to date, yearly.
5. F. NARIN, *Evaluative Bibliometrics*, Computer Horizons, Cherry Hill, N.J., 1976.
6. D. SCHWARTZMAN, *Innovation in the Pharmaceutical Industry*, Johns Hopkins University Press, Baltimore, 1976.
7. W. McVICKER, *New Drug Development Study – Final Report*, Internal Document of the U.S. Food and Drug Administration, March 1972.
8. P. de HAEN, *New Products Parade – Annual Review of the New Drugs*, de Haen Inc., New York, 1954 to date, yearly.
9. M. WINDHOLE, Ed., *The Merck Index: An Encyclopedia of Chemicals and Drugs*, Ninth Edition, Merck & Co. Inc., Rahway, N.J., 1976.
10. *Innovative Activity in the Pharmaceutical Industry*, National Economic Development Office, London, England, HMSO, 1973.
11. N. H. NIE, et al, *SPSS Statistical Package for the Social Sciences*, Second Edition McGraw Hill, New York, 1975, p. 353.
12. H. G. SMALL, "Co-Citation in the Scientific Literature: A New Measure of the Relationship Between Two Documents", *Journal of the American Society for Information Science*, 24 No. 4, August (1973) 265–269.
13. B. C. GRIFFITH, H. G. SMALL, J. S. STONEHILL, S. DYE, The Structure of Scientific Literature. 2. Toward a Macro- and Microstructure for Science *Science Studies* 4 (1974) 339–365.