



1986

The Pitfalls of Empirical Research: Studying Faculty Publication Studies

David H. Kaye
Penn State Law

Ira Mark Ellman
Arizona State University

Follow this and additional works at: http://elibrary.law.psu.edu/fac_works



Part of the [Legal Education Commons](#)

Recommended Citation

David H. Kaye and Ira Mark Ellman, *The Pitfalls of Empirical Research: Studying Faculty Publication Studies*, 36 *J. Legal Educ.* 24 (1986).

This Article is brought to you for free and open access by the Faculty Works at Penn State Law eLibrary. It has been accepted for inclusion in Journal Articles by an authorized administrator of Penn State Law eLibrary. For more information, please contact ram6023@psu.edu.

The Pitfalls of Empirical Research: Studying Faculty Publication Studies

David H. Kaye & Ira Mark Ellman

Studies of the rates at which faculty at various schools publish are like popularity contests. The winners think that these studies mean something, while the losers insist that the studies are flawed or misguided. As with any empirical research, however, there are some generally accepted criteria against which to judge a study's design and execution. Since studies of law faculty productivity are so good at commanding the attention of the legal profession, they may provide an engaging context for thinking about the nature of good empirical research.

The study by Professors Swygert and Gozansky recently published in this *Journal*¹ provides just such an opportunity. Indeed, this study promises to examine not merely relative publication rates, but to illuminate the relationship between tenure and the rate at which faculty publish intellectually meaningful scholarship. The opening paragraphs of the article state its principal purpose to be the investigation of the productivity of senior faculty whose research efforts are not motivated by concerns of "job security or promotion."² The concluding paragraphs report that 44 percent of the senior faculty published nothing during the three year study period, evidencing "an underutilization of intellectual resources."³ Although never fully articulated, the implied finding is that otherwise able scholars cease producing once the pressures of tenure and promotion have been lifted.⁴

David H. Kaye is Professor of Law and Director of the Center for the Study of Law, Science and Technology, Arizona State University.

Ira Mark Ellman is Professor of Law, Arizona State University.

1. Michael I. Swygert & Nathaniel E. Gozansky, *Senior Faculty Publication Study: Comparisons of Law School Productivity*, 35 *J. Legal Educ.* 373 (1985).
2. After observing that "an untenured law teacher's publications may be considered both predictable and coerced," *id.* at 374, Swygert and Gozansky ask: "what about senior law faculty whose scholarly pursuits presumably are not influenced by job security and promotion?" *Id.*
3. *Id.* at 393.
4. A story in the *New York Times* describes the authors as conceding that "professors at the nation's 169 law schools are productive to a point: namely, the point they receive tenure," but concluding that "with some exceptions, it's all downhill after that." *New York Times*, December 22, 1985, p. 31 at col. 1. The article quotes Professor Swygert as saying "Tenure has a down-side. It seems to me that tenure has gone beyond the protection of free speech, and has allowed a certain amount of incompetency to be protected as well." *Id.*

© 1986 by the Association of American Law Schools. Cite as 35 *J. Legal Educ.* 24 (1986).

Swygert and Gozansky may be right. With tenure or promotion attained, many faculty, including those who can produce good scholarship, may move on to other, though not necessarily trivial, pursuits merely because they have lost the incentive to achieve tenure or be promoted. The research of Swygert and Gozansky, however, does little or nothing to establish this proposition. Inasmuch as Swygert and Gozansky call for further empirical study of faculty publication rates,⁵ and to the extent that law-trained scholars are notoriously weak at empirical research,⁶ it may be worthwhile to identify some of the methodological considerations that should inform such research. These fall into four broad categories: (1) problems of conceptualization, (2) problems of measurement, (3) problems of data presentation and analysis, and (4) problems of inference. We examine each of these, drawing on the Swygert and Gozansky article for illustrative purposes.

I. Conceptualization

A. An Experimental Paradigm

Research into faculty productivity, like any worthwhile research, should begin with a clear statement of the questions to be answered.⁷ We have indicated that a major question raised by Swygert and Gozansky is whether the granting of tenure and promotion causes faculty to produce less scholarship. They apparently hypothesize that it does.

In conceptualizing how best to examine such an hypothesis empirically, it is helpful to begin by imagining how we might proceed experimentally in an ideal world of no practical constraints. Designing such a thought experiment will give us a better understanding of what should be measured and how such measurements should be interpreted. In this case, for instance, we could assign beginning law faculty at random to two groups: a "control group" that suffers the normal tribulations of the tenure process, and a "treatment group" that is hired with tenure. If the control group published more in the early years (during which only the less productive treatment group was tenured), Swygert and Gozansky's hypothesis would be corroborated. If the productivity of the two groups were equal, or if the tenured treatment group actually published more, the experiment would cast doubt on their hypothesis. It is important to recognize that the pattern of publication over time *within* a group would not strengthen or weaken the hypothesis; only the comparison between the groups would be truly probative. This is because many factors that change over time may affect publication rates. For example, publication rates may fall as a consequence of the increasing responsibilities and opportunities that come with

5. Swygert & Gozansky, *supra* note 1, at 394. One may wonder whether such research is of very high priority.

6. Of course there are notable exceptions, and considerable progress has been made since the pioneering days of Underhill Moore and his colleagues.

7. Such questions may be preliminary, and the research itself may suggest new avenues of investigation.

professional maturity. It is also conceivable that publication rates might increase with enhanced expertise, or that a variety of conflicting factors might interact to produce a constant rate of publication over time. The point is simply that many things that change over time may influence publication rates, and that we may have no way of isolating the causal influence of any one of them, including the granting of tenure and promotion. Our thought experiment avoids this difficulty since such potentially confounding factors, whatever they are, should be about equally present in our two randomly selected groups, leaving the difference in their tenure status as the principal determinant of any difference in the publication rates.

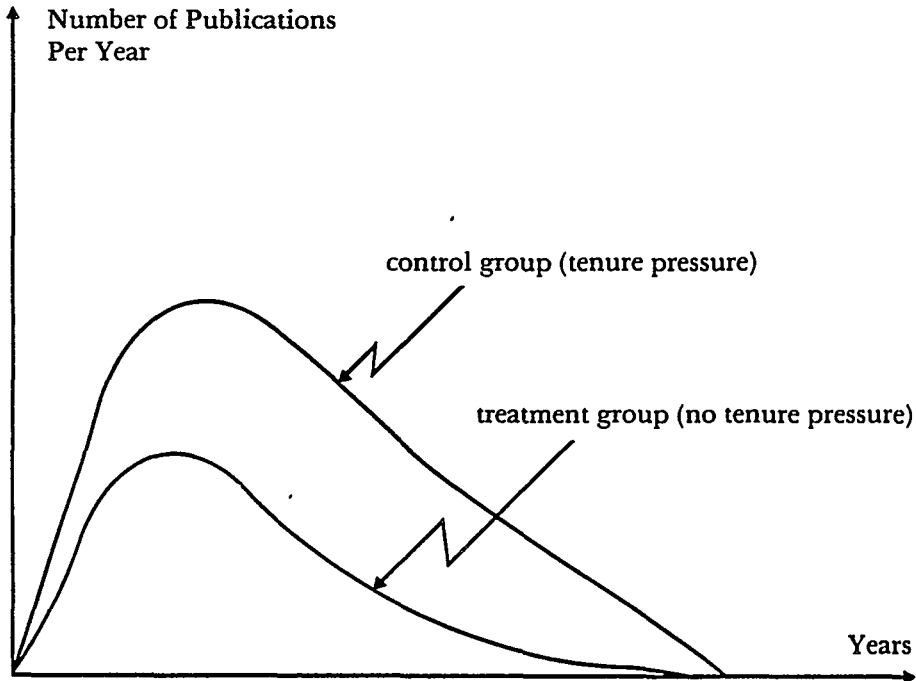


Figure One. Hypothetical Publication Rates In Controlled Experiment

The possibilities can be seen most easily by graphing hypothetical results for our hypothetical experiment. Let us assume that the results confirmed Swygert and Gozansky's hypothesis. As the prior discussion should make clear, that assumption does not require that our graph contain curves of any particular shape; it requires only that the curve for the control group be higher than the one for the experimental group. Figure One depicts such hypothetical results.

Such a controlled experiment is obviously not feasible. As a consequence, we are obliged to substitute a more fallible observational study. The most obvious way to proceed is with a longitudinal study—one that tracks productivity over the professional lifetime, or at least over an extended

period which includes both pre- and post-tenure years, of each law teacher. Yet for the reasons just reviewed, such a study cannot prove that receipt of tenure makes law teachers less productive, since even if it showed declining productivity after tenure, we would not be able to tell whether this decline was the result of tenure itself or of the other possible confounding factors that may also be associated with increasing professional age. The absence of an experimental group makes it impossible to measure the influence of any single factor. If we graphed the results of such a longitudinal study, we might get the curve for the control group in Figure One, but we would have no other curve with which to compare it.

Nonetheless, such a longitudinal study would give us some information. It would at least establish whether publication rates change over time, even if it did not help us understand whether the granting of tenure played a role in that change. A longitudinal design would be better than the design actually used by Swygert and Gozansky. As we shall see, their design cannot even tell us if there is a change in publication rates, much less what factors might account for it.

Swygert and Gozansky employ a cross-sectional study design, which is an additional step removed from the experimental paradigm, and less revealing than the longitudinal design. They look only at the publication rates of "senior faculty"—defined as those who attained the rank of full professor before 1976—during the years 1980–1982. They do not examine the publication rates of these faculty in their earlier years. That is, they examine only one slice—one "cross section"—of a group of much longer careers. They can report on the publication activity of persons caught in this particular slice, but they have no data from other points in these faculty members' careers with which to compare this particular slice. They have a picture of one point in time rather than a series of pictures capturing publication activities through time. If we were to graph their data, we would have only a few points in the tail of the control group curve in Figure One. We would not even have one complete curve, much less the two that we need to reach any causal conclusions.

If we assume that Swygert and Gozansky's measures of scholarship are sensible (more on this later), that their definition of senior faculty is not too arbitrary, and that the 1980–1982 period is representative, then we can conclude that 44 percent of senior faculty do not publish. We do not know how many of these same faculty did not publish in their junior years. Many, if not most, of the 150 or so accredited law schools⁸ had no serious scholarship requirement ten or twenty years ago when the tenure decisions were made in Swygert and Gozansky's sample group. One might thus plausibly surmise that most of the nonpublishing senior faculty also produced little before they were tenured. Or one might surmise that the bulk

8. The 1985–86 Prelaw Handbook, published by the Law School Admission Council in cooperation with the ABA and the AALS, lists 174 ABA approved law schools. LSAC, 1985–86 Prelaw Handbook 6 (1985). Ten or twenty years ago the number was obviously smaller.

of these senior nonproducers were once highly productive scholars. The point is that Swygert and Gozansky's methodology is incapable of providing information that would help in deciding which of these hypotheses is correct.

To summarize, a controlled experiment, if feasible, could answer the causal question that Swygert and Gozansky asked. A longitudinal study could show that scholarly productivity declines over time, but because many different potentially relevant but unmeasured variables also change over time, it could not identify any causal link between any of them and this observed decline. Finally, a cross-sectional design, like the one used by Swygert and Gozansky, cannot identify the existence of changes over time, and certainly cannot show any causal effect of tenure or promotion on productivity. Such a design is therefore not an appropriate vehicle for examining the relationship between the variables that Swygert and Gozansky discuss.

B. Defining the Variables

With a clearly specified research question, one is in a position to decide which variables to measure. For example, someone interested in investigating the effect of tenure might want to look at total professional activity, including things such as serving on committees of professional organizations, editing professional journals, serving on governmental agencies or commissions, or performing administrative tasks within the university. These items would be appropriate in establishing whether the granting of tenure resulted in lazy faculty who abandoned professional work. As a second possibility, one might want to look only at a subset of professional activities, such as the production of professional writing—that is, writing on legal or educational topics. A third possibility would be to look only at scholarly writing.⁹ This might be appropriate if one started with the premise that because scholarship is what is required for tenure, many faculty will only produce scholarship when under pressure induced by an impending tenure decision.

The Swygert and Gozansky study looks to the gross number of publications in a given period. As the researchers recognize, this count of

9. Scholarly writing and professional writing are not always the same. There is much useful professional writing that is not scholarship under most conventional definitions. Examples include pure descriptions of new statutes or cases, or material prepared exclusively for teaching purposes, which limits itself to a carefully sequenced presentation of well-known principles. At the same time, it can be argued that much scholarship is not professional writing, if one defines professional writing as that which yields insights of value in the practice of law. We offer no judgments on whether professional nonscholarship, or scholarly nonprofessional writing, truly exist, or whether either is valuable, or should be counted toward tenure. Our only purpose is to indicate that such distinctions may be important to someone who sets out to measure something, because the method of measurement may not be appropriate depending upon whether one wants to include both kinds of writing or just one.

publications does not measure the full range of professional activities.¹⁰ Their references to "research" and their initial hypotheses concerning the impact of tenure pressures suggest that they would want to measure the production of scholarly writing rather than all professional writing.¹¹ Yet the pool of material from which they measure the gross number of publications is the more than 700 items referenced in the Legal Resources Index. This index covers not merely the full gamut of law reviews, from highest to lowest quality, but also local and state bar journals, legal newspapers, and various publications serving practitioners, such as the *Medical Trial Techniques Quarterly*, the *Lawyers Title Guarranty Funds News*, and the *Licensing Law and Business Report*. Such a method obviously counts all professional writing, rather than legal scholarship per se.

It is not necessarily wrong to count all professional writing, but the authors of an empirical study must decide what variables they want to measure, design a method of measurement appropriate to that decision, and discuss the significance of their results in a manner that takes the measurement technique into account. The measurement method employed by Swygert and Gozansky appears to be a poor indicator of professional activity in general, or scholarly activity in particular, and the authors never discuss why one would want to measure the intermediate level of professional writing instead of these other two levels.

Of course, Swygert and Gozansky may have had no desire to examine the relationship between tenure and scholarship or between tenure and professional activity. Their remarks about tenure and the differences between junior and senior faculty may be the equivalent of *obiter dicta*, in that all they really meant to describe was the publishing productivity of senior faculty in law journals generally,¹² or for the sake of making comparisons among institutions. If so, the study is more like the usual popularity contest after all. Nevertheless, these contests are often provocative in the status conscious world of academia, and, as with more serious research, there are good and bad ways to stage them. The remaining sections of this Comment therefore consider whether the Swygert and

10. Swygert and Gozansky acknowledge that the "nonproducers" they identify may have "teaching loads, committee work, or other institutional service that constrained opportunities for research." Swygert & Gozansky, *supra* note 1, at 393. Although this list omits some plausible professional activities, Swygert and Gozansky presumably would not discount their potential impact. Yet they conclude that the intellectual resources of the nonproducers are "underutilized." Evidently, they believe that either (1) there is a very strong correlation between publishing and other professional activities, or (2) anyone who does not write for publication is underutilizing his or her intellectual resources.

11. Indeed, their article appears in a Symposium entitled "Law Professors, Lawyers, and Legal Scholarship."

12. One might well ask why anyone would want to collect such data. If it is because senior faculty have fewer external pressures on them to produce, as Swygert and Gozansky emphasize, then one is not interested in the data on senior faculty for its own sake, but only to the degree that it speaks to the relationship between productivity and tenure or promotion. In that event, as we have seen, one needs a very different research design.

Gozansky study gives much insight into the relative written productivity of senior faculty at various institutions. It will come as no surprise by this point that we have our doubts.

II. Measurement

We have thus far focused upon the problem of identifying what one wants to measure. After these variables have been defined, one must select a suitable method for measuring them. This measurement method should count things in a sensible way and should accurately measure the variables.

To raise the question of what variables should be measured in a faculty productivity study, we previously observed that Swygert and Gozansky count local bar journal or newsletter writing along with books and law review writing. To illustrate the difficulties that can arise in operationalizing the process of measuring even an apparently ill-chosen variable such as total published output, we now note that Swygert and Gozansky count every piece of writing equally. As they explain, "[a]ll publication items that come within the inclusion criteria are treated as statistically equal events regardless of length, topic, documentation, originality, impact, or journal of publication."¹³ Laurence Tribe's award-winning *American Constitutional Law*, a five page book review in an ABA Section publication, a newsletter note on a recent conference, and a county bar journal note on the local domestic relations court's new alimony schedule each would count as an equivalent item. Even if one merely seeks to measure total output, as Swygert and Gozansky apparently do, this seems an odd way to proceed.

Swygert and Gozansky's measurement methods are especially puzzling because alternative measures are readily available. One can count pages as well as items,¹⁴ one can develop criteria to distinguish between scholarship and journalism,¹⁵ one can count publications in scholarly journals, or a subset of scholarly journals, differently or separately from other publications,¹⁶ and one can present alternative measures in the same study,

13. Swygert & Gozansky, *supra* note 1, at 375.

14. See Ira M. Ellman, A Comparison of Law Faculty Production in Leading Law Reviews, 33 J. Legal Educ. 681 (1983).

15. Cf. Fred R. Shapiro, The Most-Cited Law Review Articles, 73 Calif. L. Rev. 1540 (1985).

16. See Ellman, *supra* note 14. Swygert and Gozansky evidently rejected this approach on the ground that ascertaining which journals are scholarly or which of the scholarly journals tend to publish the better articles entails too much subjectivity. They write that "[t]he Ellman study . . . arguably comes closer to measuring quality of publications by limiting its scope to articles placed in what Ellman considered to be the top twenty-three law journals in America." Swygert & Gozansky, *supra* note 1, at 364 n. 66. It should be obvious, however, that objective criteria exist to permit more focused measurement. Besides discarding, or reporting separately, publications in certain types of journals, one can do at least a crude content analysis to distinguish between descriptive and analytical writing. Finally, one can identify specific journals that tend to publish more highly regarded scholarship without resort to the researcher's personal opinion. Indeed, contrary to what Swygert and Gozansky suggest, the Ellman study presents one method of doing exactly this, and this method was not based on any personal judgment as to the best law journals. That study instead identified certain journals by reference to two objective criteria: the frequency with which they were cited, and their affiliation with law schools that ranked high in reputational studies conducted by other researchers. See Ellman, *supra* note 14. Many variations on this approach to categorizing publications are possible.

to allow the reader to see the extent to which the results of the study depend upon the measurement methods used.¹⁷

III. Analysis and Presentation

Once one has obtained valid measurements of the appropriate quantities, one must decide how to present them. This ordinarily requires some analysis of the data to render it more compact and more revealing. While the data may be transformed, if only by reordering, a good presentation does not deny the reader the information necessary to draw his or her own conclusions about the meaning and importance of the results. To the contrary, the author should facilitate such an effort by presenting the data in a convenient form which makes its meaning clear while striving to present all the information needed to reach the conclusions that the data suggest. In sum, while the text typically characterizes or speculates upon the data, accompanying tables and figures should present the numbers needed to allow the intelligent reader to make an independent evaluation of the results.

There are many useful graphical tools for revealing and communicating the meaning of the data. In addition to traditional graphical displays such as histograms, the newer techniques of exploratory data analysis such as stem and leaf diagrams and box plots¹⁸ can be very helpful in this phase of the research. It is unfortunate that so much of the empirical research appearing in law journals makes so little use of these techniques. Swygert and Gozansky's article is a case in point. Replete with lists, it contains not a single picture to aid the reader in making sense of the tables of numbers.¹⁹

Worse yet, many of the tables discard important information for no apparent reason, making it impossible for a statistically literate reader either to make an independent assessment of the significance of the reported conclusions or to organize the data in a more revealing way. For example, Swygert and Gozansky frequently tabulate schools by groupings and rankings but fail to report the totals associated with each school from which they generated these rank-order statistics. Nowhere in the article do they report the number of items produced by the named schools, the number of senior faculty at these schools, or the mean and variance (or more complete statistics) for the number of items published per senior faculty member at the various schools.

17. Ellman, *supra* note 14, gives both an item and a page count. It is conceivable that Swygert and Gozansky's results are robust—that the use of other measures would not change the picture much. This would be the case if the type of writing (descriptive versus analytical, scholarly versus journalistic, etc.) and the place of publication were uncorrelated with the law schools in which faculty teach. Our impression is that this condition does not hold, and Swygert and Gozansky present no data to the contrary.

18. E.g., John M. Chambers, William S. Cleveland, Beat Kleiner & Paul A. Tukey, *Graphical Methods for Data Analysis* (Boston, 1983).

19. Swygert and Gozansky state that they did undertake some graphical analysis, but they decided not to present it "to avoid confusion." Swygert & Gozansky, *supra* note 1, at 392 n. 59.

IV. Inference

In addition to describing the data, an empirical study should usually include information to indicate the margin of statistical error in some of the descriptive statistics.²⁰ Typically, there is some randomness in the process that generates the data, so that another data set produced by the same process would not be identical. If the randomness is large, then the descriptive statistics, although properly computed, may not be trustworthy. One useful method to reveal the likely extent of statistical error is to present these statistics as interval estimates. For instance, rather than merely giving a sample mean as a point estimate of a population mean, one can provide an interval estimate—a range of plausible values for the population mean.²¹ Similarly, rather than merely reporting a correlation coefficient, one can provide an interval estimate for that quantity or one that is of more direct interest.

These abstractions can be clarified by turning to the Swygert and Gozansky study. The authors assert that “there are significant differences in the frequency and extent of senior law faculty publishing among the 138 AALS schools.”²² It is difficult to know whether the term “significant” here is used in its technical sense—to indicate that certain differences would not occur very often if they were simply the consequence of the randomness in the process that generated the statistics—or in its ordinary language sense of “large” or “substantial.” Had the authors actually stated the magnitude of the differences that they had in mind and given an interval estimate for these differences, there would have been no ambiguity, and the reader could have come to his or her own conclusion about their “significance.”

Likewise, Swygert and Gozansky’s failure to use interval estimates or some comparable technique mars their study of the correlation between the

20. There are problems in interpreting significance probabilities when the data set is not a random sample from a population. See, e.g., Persi Diaconis, *Theories of Data Analysis: From Magical Thinking through Classical Statistics*, in *Exploring Data Tables, Trends, and Shapes*, eds. David E. Hoaglin, Frederick Mosteller & John W. Tukey, 1, 27 (New York, 1985). Thus, some statisticians would argue that any use of inferential statistics with the Swygert and Gozansky data is misguided.

21. For a more careful and complete statement of the meaning of an interval estimate, see, e.g., David H. Kaye, *Is Proof of Statistical Significance Relevant*, Wash. L. Rev. 1986 (in press), available as Research Report No. 86-1, Center for the Study of Law, Science and Technology, Arizona State University.

We have already remarked that in the Swygert and Gozansky paper, even simple, descriptive means for publication rates by school are lacking. Consequently, we would not expect to find interval estimates of the means for various schools. Although Swygert and Gozansky, *supra* note 1, at 393, refer to something they call “a standard deviation analysis,” which might be pertinent here, it is impossible to tell what they analyzed, how they analyzed it, or why. The courts also are fond of this same phrase. See, e.g., David H. Kaye, *Hypergeometric Confusion in the Fourth Circuit*, 26 *Jurimetrics J.* 1986 (in press), but unlike the standard deviation itself, the phrase “standard deviation analysis” has no conventional meaning in statistics.

22. Swygert & Gozansky, *supra* note 1, at 391.

size of a school's senior faculty and the publication rate of that faculty.²³ They claim that there is "a moderately strong relationship" between senior faculty size and various measures of school productivity.²⁴ In reaching this conclusion, they compute correlation coefficients of about .5 between size and different measures of productivity. They assert that "anything over about .2 indicates a perceptible relationship," and that "[t]he larger the study population, . . . the more reliable the correlation coefficient becomes as a predictor of performance."²⁵ Because their study includes 169 AALS member schools,²⁶ they conclude that they have demonstrated "a substantial relationship" between senior faculty size and productivity.²⁷

Once again, this hand-waving analysis confuses statistical significance and substantiality. Psychologists often use the correlation coefficient to describe the strength of an association between two variables, although better statistics are usually available.²⁸ In interpreting a correlation

23. The point of this enterprise is a bit mysterious. If it is to generate some useful hypotheses about what factors influence senior faculty productivity, it is disappointing that Swygert and Gozansky do not describe such hypotheses. It seems doubtful that size per se is a causal factor, but the causal factors that it might be acting as a proxy for are left unmentioned.

24. Swygert & Gozansky, *supra* note 1, at 392.

25. *Id.*

26. Swygert and Gozansky did not reveal the correlation for the nonmember schools. They assert that the correlation is higher for these schools because the coefficient for all ABA schools (a group that includes all AALS schools) is slightly higher (by an amount of .012 or .59, depending on the measure of productivity). They deduce that "a non-AALS school's publication rate may be more strongly affected by outside sources, such as size," which "makes it reasonable to eliminate the non-AALS schools" from the analysis. *Id.*

We find this line of reasoning difficult to fathom. At the outset, the premise—that if one data set is pooled with another data set that has a correlation of .5, then to yield a combined set with a correlation in excess of .5 the correlation within the added data set must exceed .5—is demonstrably false. See, e.g., Allen L. Edwards, *An Introduction to Linear Regression and Correlation* 56-60 (San Francisco, 1976). Second, even if the non-AALS schools do exhibit a correlation coefficient larger than that of the AALS schools, it is not clear that the difference in the coefficients represents anything more than the randomness inherent in the process that generated the productivity data. The schools not accredited by the AALS are probably much younger, on average, than the AALS schools. The sizes of the senior faculties could be smaller, and the publication rates subject to more variability. If so, an interval estimate of the confidence coefficient would be broad, making it more difficult to detect any difference in the correlations for AALS and non-AALS schools. Because Swygert and Gozansky do not present the relevant statistical analysis, it is impossible to know whether the small differences that they point to would be likely to arise under a model that recognizes randomness in the data. Finally, even if the differences are in the direction that Swygert and Gozansky presume, and even if they are unlikely to arise under a chance model (i.e., they are statistically significant), the remaining inferences are nonsequiturs. Such differences cannot show that the productivity of faculty at either group of schools is "affected by" size. The most that could be said on the basis of the data is that productivity is associated with size, and not with unnamed factors "such as size." Nor is it obvious why a stronger association in non-AALS schools would justify restricting the analysis to AALS schools. Since we are uncertain of the point of this analysis to begin with, however, we may be overlooking some consideration that Swygert and Gozansky had in mind.

27. Swygert & Gozansky, *supra* note 1, at 392.

28. Of course, the correlation coefficient is not itself a "predictor of performance," as Swygert and Gozansky infelicitously put it. A regression rather than a correlation study could give such a predictor.

coefficient, one should understand that the strength of the relationship has nothing to do with the sample size,²⁹ and there is no reason to say that a correlation coefficient of .5 is moderately strong merely because it is bigger than .2 and was computed with a large data set. In the absence of a baseline of other correlation studies to draw on, “moderately strong” is nothing but a more or less arbitrary translation of the number itself into words, with no established technical meaning. Another way to express the same fact, for instance, is to say that faculty size “explains only 25% of the variance in productivity.”³⁰

On the other hand, in determining whether the observed correlation coefficient of .5 or so could be a fluke—a result that would not be likely to recur if another data set were generated and analyzed—there are simple procedures to follow. They do not involve comparing anything to .2. An interval estimate for the true correlation coefficient would show whether the value lies in a region indicative of a nonzero correlation between senior faculty size and publication rate. Although it probably does, there is no good reason to leave this to the reader’s imagination.

* * *

We have devoted this much attention to Swygert and Gozansky’s article for several reasons. In part, we have tried to explain why we found it unpersuasive, but our principal motivation goes beyond this particular study. We have tried to inform other lawyer-researchers who care to study the influences on law faculty productivity about some of the methodological issues that they must confront. We hope that our critique will help ensure that empirical studies by attorneys are held to the minimal standards of research competence that are to be found in other academic fields which rely on empirical studies.

29. For instance, the strength of the relationship in the population between height and weight is not affected by the size of the sample surveyed. The population correlation coefficient is whatever it happens to be, and its magnitude is a measure of the strength of the linear relationship. A correlation coefficient computed on a sample of random observations from the population provides an estimate of the magnitude of the population coefficient. If the sample coefficient were .6, for instance, one might estimate the strength of the relationship in the population to be .6. This would be so regardless of the size of the sample that produced the coefficient of .6.

The size of the sample does affect the accuracy of the estimate of the correlation in the population, but that is a different matter. If the approach recommended in the text were adopted, and interval estimates rather than point estimates were given, then the width of the interval would decrease as sample size increased, but (apart from statistical fluctuations in the random sampling) the midpoint of the interval—our best guess of the population statistic—would not change.

30. Neither statistic tells us how much the mean publication rate grows as one increases faculty size a given amount. Cf. *supra* text accompanying note 21.