The Journal of Sociology & Social Welfare

Volume 16
Issue 1 *March*

Article 4

March 1989

When Is Statistical Significance Meaningful? A Practice Perspective

Robert W. Weinbach
University of South Carolina

"When Is Statistical Significance Meaningful? A Practice Perspective"

ROBERT W. WEINBACH

University of South Carolina
College of Social Work

Statistically significant relationships may be quite small in the absolute sense. The practitioner who faces the issue of when to utilize a finding for practice must consider more than mathematical arguments. The place of practice considerations in decision making for utilization is discussed.

When a report of empirical research states that the relationship between two variables was found to be statistically significant, there is a normal tendency to be impressed. After all, the researcher has scientifically and objectively demonstrated that the variables are related and has generated knowledge that can be used. For the social work practitioner, the finding can form the basis for an important behavioral change. However, a statistically significant relationship between variables may or may not be worthy of note or of implementation. The relationship is believed to be real, but it may not be very strong.

The difference between relationships that are merely statistically significant and those that are statistically significant but also are strong enough to be considered meaningful surfaced in the social work professional literature several years ago. It took the form of a heated debate between educators. Gould and Kim (1976, p. 50) reported their research findings on "the effects of sex on salary differentials" between social work faculty. In a critique of the Gould and Kim research as well as other research that examined the possibility of sexual discrimination within social work, Allen Rubin reminded journal readers that statistically significant relationships are not necessarily strong ones. He noted that: A caution all too often ignored in the social work literature is that statistically significant relationships need not be strong enough to be meaningful for practical purposes. Triv-

ial relationships can be statistically significant if based on large samples (Rubin, 1981, p. 22).

Rubin's unfortunate use of the words "trivial" and "minuscule" (p. 23) to describe the differences between salaries of male and female educators resulted in a response of outrage that tended to obscure a very important message for the social work practitioner. In a rebuttal, Gould (1983, pp. 34–35) focused most of her attention on defense of her research methods and on discrediting Rubin's. What might have been an important and useful dialogue on the distinction between "statistically significant findings" and "meaningful findings" quickly degenerated into a question of whose research methods were better and whose biases may have been showing.

The typical social work practitioner is unlikely to read research reports on a regular basis, or to use them for practice decision-making (See, for example, Rosenblatt, 1968; Kirk & Fischer, 1976). If this unfortunate condition is to change, practitioners will need to become knowledgeable in ways to critically evaluate the findings of research reports. This should entail, among other things, the assessment of whether a finding is of sufficient importance to incorporate into one's everyday decision making and professional behavior. The decision is not one that requires a high level of sophistication in either research knowledge or statistics. It does, however, require that practitioners assume a position of healthy skepticism toward statistical significance and be prepared to apply common sense in evaluating whether a statistically significant relationship between variables is meaningful for them. A review of both the concept of statistical significance and how it is achieved is helpful in making this important distinction.

The term "statistical significance" is especially problematic for the social worker with a mindset that is more within practice than within statistics. In a social work practice context, we become conditioned to think of significance as synonymous with importance. For example, we talk of "significant others" or "significant relationships," communicating the idea that a person has played an important role in the life of another. To the statistician, however, significance is quite independent of importance.

The conclusion that two variables reflect a statistically significant relationship is a mathematical determination based upon nothing more than the laws of objective probability. The computer that performs the statistical operation has no insight into the nature of the variables or of the importance in human terms of any relationship that may exist between them. It cannot know whether the degree of the relationship uncovered is of practical value to the practitioner or, if it is, just how valuable it might be. To say that the two variables reflect a statistically significant relationship to each other is little more than a statement of a reasonable assurance. In most instances it means nothing more than that researchers are comfortable in concluding that the variables are related. They have demonstrated mathematically that, in drawing this conclusion, there is an acceptably small chance of being wrong. They are reasonably certain that they will not commit a Type I error; that is, they will not conclude that the variables are related when they really are not. It is a calculated gamble that, by convention, we believe to be justified in the interest of bringing the always provisional knowledge of science to light.

Of course, bias or the presence and influence of other variables (in addition to chance) also may have caused an apparent relationship. But, even if all competing explanations were controlled or ruled out by rigorous research design and/or statistics and we are left with the conclusion that the relationship between or among variables is probably a true one, the issue of the value of the finding must still be addressed. Whether the relationship identified is one of cause and effect (resulting from a tightly controlled experiment) or the more common association or correlation resulting from less rigorous designs often seen in social work research, it still might be "no big deal," at least not one suggesting a change in a practitioner's behavior.

The issue of whether a statistically significant relationship is meaningful or trivial is one of both strength of the relationship and one of professional values and priorities. Rubin's (1981, p. 22) perception of meaningfulness focused on the former. He pointed out with compelling mathematical logic that the mean salary differential between men and women (after controlling for certain variables) was $301, "only" a difference of 1.1%. Others

might have viewed the finding differently while focusing on values. They may have believed that insufficient attention was given to the fact that the profession of social work is committed to elimination of sexism within its ranks. The $301 may not have been trivial to them; in this sense, a difference of even $10 might have been meaningful. (Mathematically, such a difference could be statistically significant, given a large enough sample). The point is, there was a difference, as demonstrated by statistical analysis. Whether this was a finding worthy of behavioral change was a matter of individual perception and opinion.

The phenomenon of statistically significant but weak relationships between variables with large samples is real, and should not be ignored. It has been demonstrated (Weinbach and Grinnell, 1987, pp. 124–125) that, for example, a crosstabulation that results in a significance level of pp $>$ .20 (not statistically significant) quickly jumps to $p < .01$ (significant) if the frequencies in all cells are multipled by ten. Similarly, large samples can result in correlations that are statistically significant, yet the correlation itself is so low as to be of little worth in its ability to predict the value of one variable from knowing the value of the other for a given case. Clearly, there is a threshold where a statistically significant relationship between variables becomes trivial for the individual social worker. But where is that threshold? Rubin and Conway (1985) suggested one possibility. They proposed a mathematical solution to the dilemma experienced by the consumer of research who is attempting to separate the meaningful from the trivial. While a step in the right direction, their proposal still suggests that the issue and its solution lie primarily in the world of statistics.

The presence of a relatively large sample size should alert the reader familiar with statistics that further inquiry is needed. It may indicate that a situation of a statistically significant but trivial relationship may exist. But it also may not. Only practitioners, not statisticians, can make the final decision as to whether the relationship between variables and the strength of the relationship are meaningful to them, i.e., suggest the need for behavioral change.

The critical evaluation of research may require more infor-

mation than is currently reported in some professional journal articles. If practitioners are to make intelligent and informed decisions regarding whether a finding is meaningful for them, they must know more than just whether a relationship between variables was statistically significant. They must also be told the sample size and the strength of the relationship in a readily comprehensible style, for example, percentage difference or actual correlation. Rubin and Conway (1985) argued for the inclusion of some indicator of relationship strength. They recommended that "researchers routinely report and interpret a magnitude-of-relationship statistic of every statistically significant relationship." To save the reader the time of looking up a description of the appropriate usage of a statistical test, a statement of whether the sample size fell within the usual size range for which the test is best suited should also be included. If not, some explanation of why the test was used is probably indicated. The ethical researcher who invites replication and feels comfortable in use of statistical testing should not object to any of these requirements. Editors may require explanation in order to recognize that these details are anything but superfluous.

Given sufficient information about the research sample and the results of statistical tests, how do practitioners decide when a statistically significant finding is sufficiently meaningful for their practice utilization? The decision is an individual one based on considerations that include economics, time considerations, professional judgement and other factors inherent within the practice environment. A statistically significant finding may be trivial within Agency X, but substantive within Agency Y.

A hypothetical research finding may help to illustrate the point. Suppose a research report or article based on a study using a relatively large sample were to report that a new treatment (B) was associated with a higher level of self concept among depressed adolescents than was the usual treatment (A). The relationship between the dependent and independent variables was statistically significant with a mean score of 79 for B on a standardized self-concept index as opposed to a mean score of 75 for treatment A on the same index. Readers of the report are sufficiently impressed with the rigor of the design. They are convinced that the relationship reported is a real one and that

the magnitude-of-relationship between variables is sufficiently large. However, the finding may still be considered trivial to them. Why? If the hypothetical readers are administrators, they must consider the cost and consequences of utilizing the finding. After some thought they conclude that:

1. A four point difference is not really much in the absolute sense (on a self concept index with a range of 100).
2. It would use most of the current annual continuing education budget to provide assistance to professional staff to "retool" to be able to use treatment B.
3. Some key and valued staff members may have a heavy investment of time and reputation in using Treatment A and would resist using Treatment B. They might, e.g., become fearful over loss of status if Treatment A is no longer used, sabotage implementation of Treatment B or even quit their jobs.
4. Within the agency, problems of effectiveness are far less severe in treatment for low self-concept than they are for treatment of other problems. Improvement of effectiveness in this area of treatment is a low priority.

The decision that the statistically significant relationship was not meaningful was based only in part on the statistical report of findings. The final determination was made on data that were derived from insight into the current work environment and from common sense and practice logic. Other practicioners in another practice environment might consider the findings and decide that they are meaningful and therefore, worthy of implementation. Perhaps, these practitioners can afford the continuing education cost. They may also have a staff that is actively seeking effective treatment methods for addressing a perceived severe problem of ineffectiveness in working with clients who possess a low self-concept.

Both hypothetical practitioners used the same finding for decision making, but they arrived at different decisions. They were aware that relationships between variables can easily be statistically significant with large samples. They made an informed judgement as to whether or not the relationship was meaningful for their practice needs. The decision not to implement the findings was as sound as the one to implement; both were empirically based, but both also took into consideration the world of practice.

The reminder that statistically significant relationships may be trivial in the absolute sense is, on one level, a small but useful piece of information for the practitioner who aspires to utilize research. But a second reminder that professional judgement is essential to decision-making regarding research utilization may be of greater importance. Research utilization occurs best when it entails a practical mix of knowledge of the scientific method and sound practice judgement. In an environment that fosters research utilization, researchers, sensitive to the decision-making needs of practitioners, will conduct research and communicate findings in a way that they can be used by practitioners. Practitioners will be more likely to read research reports and to utilize them when they can see the place of practice knowledge in the utilization process.

References

Gould, K. and Kim, B. (1976). Salary inequities between men and women in schools of social work: Myth or reality? *Journal of Education for Social Work, 12,* 50–55.

Gould, K. (1983). Another look at salary inequities. *Social Work Research and Abstracts, 19,* 34–35.

Kirk, S. and Fischer, J. (1976). Do social workers understand research? *Journal of Education for Social Work, 12,* 63–70.

Rosenblatt, A. (1968). The practitioner's use and evaluation of research. *Social Work, 13,* 53–59.

Rubin, A. (1981). Reexamining the impact of sex on salary: The limits of statistical significance. *Social Work Research and Abstracts, 17,* 22.

Rubin, A. and Conway, P. (1985). Standards for determining the magnitude of relationships in social work research. *Social Work Research and Abstracts, 21,* 39.

Weinbach, R. and Grinnell, R. (1987). *Statistics for social workers.* New York: Longman, Inc.