Comment

Contents

How Are We Doing in Soft Psychology?

Robert Rosenthal
Harvard University

This comment is addressed to those of us who work in the softer, wilder areas of our field—the areas in which the results seem ephemeral and unreplicable and in which the r^2s seem always to be approaching zero as a limit. These softer, wilder areas include those of clinical, developmental, educational, organizational, personality, social, and health psychology. They also include parts of psychobiology and cognitive psychology. My message to those of us who toil in these muddy vineyards is that we are doing better than we might have thought we were doing.

How Large Must an Effect Be To Be Important?

There is some good news and some bad news abroad. The good news is that more sophisticated editors, referees, and researchers are becoming aware that reporting the results of a significance test is not a sufficiently enlightening procedure to stand alone. More and more we are beginning to see a report of the magnitude of the effect accompanying the p level. The bad news is that we are still not quite sure what to do with such a report of the magnitude of the effect—for example, a correlation coefficient.

There is one bit of training that all psychologists have undergone. From undergraduate days on we have been taught that there is only one proper thing to do when we see a correlation coefficient: We must square it. For most of the softer, wilder areas of psychology, squaring the correlation coefficient tends to make it go away—vanish into nothingness as it were. That is one of the sources of malaise in the social and behavioral sciences. It is sad and quite unnecessary, as we will soon see.

The Physician's Aspirin Study

At a special meeting held December 18, 1987, it was decided to end prematurely a randomized double blind experiment on the effects of aspirin on reducing heart attacks (Steering Committee of the Physician's Health Study Research Group, 1988). The reason for the unusual termination of this experiment was that it had become so clear that aspirin prevented heart attacks (and deaths from heart attacks) that it would be unethical to continue to give half of the physician research subjects a placebo. Now what do you suppose was the magnitude of the experimental effect that was so dramatic as to call for the termination of this research? Was r^2 .90, or .80, or .70, or .60, so that the corresponding rs would have been .95, .89, .84, or .77? No. Well, was r^2 50, .40, .30, or even .20, so that the corresponding rs would have been .71, .63, .55, or .45? No. Actually, what r^2 was, was .0011, with a corresponding r of .034.

Table 1 shows the results of the aspirin study in terms of raw counts, percentages, and as a Binomial Effect Size Display (BESD). This display is a way of showing the practical importance of any effect indexed by a correlation coefficient. The correlation is shown to be the simple difference in outcome rates between the experimental and the control groups in this standard table that always adds up to column totals of 100 and row totals of 100 (Rosenthal & Rubin, 1982).

This type of result seen in the physicians' aspirin study is not at all unusual in biomedical research. Some years earlier, on October 29, 1981, the National Heart, Lung, and Blood Institute discontinued its placebo-controlled study of propranolol because results were so favorable to the treatment that it would be unethical to continue to withhold the life-saving drug from the control patients. What was the magnitude of this effect? Once again, the effect size r was .04 and the leading digits of the r^2 were .00! Behavioral researchers are not used to thinking of rs of .04 as reflecting effect sizes of practical importance. But when we think of an r of .04 as reflecting a 4% decrease in heart attacks, the interpretation given r in a BESD, the r does not appear to be quite so small, especially if we can count ourselves among the 4 per 100 who manage to survive (Rosenthal, 1984).

Table 1
Effects of Aspirin on Heart Attacks Among 22,000 Physicians

Measure	Heart attack	No heart attack	Total
Raw counts			
Aspirin	104	10,933	11,037
Placebo	189	10,845	11,034
Total	293	21,778	22,071
Percentages			
Aspirin	0.94	99.06	100
Placebo	1.71	98.29	100
Total	1.33	98.67	200
Binomial effect size display			
Aspirin	48.3	51.7	100
Placebo	51.7	48.3	100
Total	100	100	200

Table 2
Other Examples of Binomial Effect Size Displays

Measure	Variable		Total
Vietnam service and alcohol problems (r = .07)			
	Problem	No problem	
Vietnam veteran	53.5	46.5	100
Non-Vietnam veteran	46.5	53.5	100
Total	100	100	200
AZT in the treatment of AIDS (r = .23)			
	Death	Survival	
AZT	38.5	61.5	100
Placebo	61.5	38.5	100
Total	100	100	200
Benefits of psychotherapy (r = .32)[a]			
	Less benefit	Greater benefit	
Psychotherapy	34	66	100
Control	66	34	100
Total	100	100	200

Note. AZT = azidothymidine; AIDS = acquired immune deficiency syndrome.
[a] The analogous r for 345 studies of interpersonal expectancy effects was essentially the same (Rosenthal & Rubin, 1978).

Additional Results

Table 2 gives three further examples of BESDs. In a recent study of 4,462 Army veterans of the Vietnam War era (1965–1971), the correlation between having served in Vietnam (rather than elsewhere) and having suffered from alcohol abuse or dependence was .07 (Centers for Disease Control, 1988). The top display of Table 2 shows that the difference between the problem rates of 53.5 and 46.5 per 100 is equal to the correlation coefficient of .07.

The center display of Table 2 shows the results of a study of the effects of azidothymidine (AZT) on the survival of 282 patients suffering from acquired immune deficiency syndrome (AIDS) or AIDs-related complex (ARC; Barnes, 1986). This result of a correlation of .23 between survival and receiving AZT (an r^2 of .054) was so dramatic as to lead to the premature termination of the clinical trial on the ethical grounds that it would be improper to continue to give placebo to the control group patients.

As a footnote to this display let me add the result of a small, informal poll I took recently of some physicians spending the year at the Center for Advanced Study in the Behavioral Sciences. I asked them to tell me of some medical breakthrough that was of very great practical importance. Their consensus was that the breakthrough was the effect of cyclosporine in increasing the probability that the body would not reject an organ transplant and that the recipient patient would not die. A multicenter randomized experiment was published in 1983 (Canadian Multicentre Transplant Study Group, 1983). The results of this breakthrough experiment were less dramatic than the results of the AZT study. For the dependent variable of organ rejection, the effect size r was .19 (r^2 = .036); for the dependent variable of patient survival, the effect size r was .15 (r^2 = .022).

The bottom display of Table 2 shows the results of a famous meta-analysis of psychotherapy outcome studies reported by Smith and Glass (1977). An eminent critic believed that the results of their analysis sounded the death knell for psychotherapy because of the modest size of the effect. This modest effect size was an r of .32 accounting for "only 10% of the variance."

Examination of the bottom display of Table 2 shows that it is not very realistic to label as *modest indeed* an effect size equivalent to increasing a success rate from 34% to 66% (e.g., reducing a death rate or a failure rate from 66% to 34%). Indeed, as we have seen, the dramatic effects of AZT were substantially smaller (r = .23), and the "breakthrough" effects of cyclosporine were smaller still (r = .19).

Telling How Well We're Doing

The Binomial Effect Size Display is a useful way to display the practical magnitude of an effect size regardless of whether the dependent variable is dichotomous or continuous (Rosenthal & Rubin, 1982). An especially useful feature of the display is how easily we can go from the display to an r (just take the difference between the success rates of the experimental vs. the control group) and how easily we can go from an effect size r to the display (just compute the treatment success rate as .50 plus one half of r and the control success rate as .50 minus one half of r).

One effect of the standard use of a display procedure such as the Binomial Effect Size Display to index the practical value of our research results would be to give us more useful and more realistic assessments of how well we are really doing as researchers in the social and behavioral sciences. Use of the BESD has, in fact, shown that we are doing considerably better in our softer, wilder sciences than we may have thought we were doing. It would help keep us better apprised of how we are doing in our sciences if we routinely translated the typical answers to our research questions to effect sizes such as r (and to its equivalent displays) and compared them with other well-established findings such as those shown in Tables 1 and 2.

REFERENCES

Barnes, D. M. (1986). Promising results halt trial of anti-AIDS drug. *Science, 234,* 15–16.

Canadian Multicentre Transplant Study Group. (1983). A randomized clinical trial of cyclosporine in cadaveric renal transplantation. *New England Journal of Medicine, 309,* 809–815.

Centers for Disease Control Vietnam Experience Study. (1988). Health status of Vietnam veterans: 1. Psychosocial characteristics. *Journal of the American Medical Association, 259,* 2701–2707.

Rosenthal, R. (1984). *Meta-analytic procedures for social research.* Beverly Hills, CA: Sage.

Rosenthal, R., & Rubin, D. B. (1978). Interpersonal expectancy effects: The first 345 studies. *The Behavioral and Brain Sciences, 3,* 377–386.

Rosenthal, R., & Rubin, D. B. (1982). A simple, general purpose display of magnitude of experimental effect. *Journal of Educational Psychology, 74,* 166–169.

Smith, M. L., & Glass, G. V. (1977). Meta-analysis of psychotherapy outcome studies. *American Psychologist, 32,* 752–760.

Steering Committee of the Physicians Health Study Research Group. (1988). Preliminary report: Findings from the aspirin component of the ongoing physicians' health study. *The New England Journal of Medicine, 318,* 262–264.

A Reply to Page: Fraud, Pornography, and the Meese Commission

Douglas E. Mould
Wichita, KS

Page's (March 1989) endorsement of the Meese pornography commission (Department of Justice, 1986) and his attack on pornography researchers is ill conceived; the contention that "it is possible to justify severe legal restrictions on the basis of present evidence" (p. 580) is quite dangerous—and wrong. Whereas Page made much of the scientific literature on pornography, he totally ignored mentioning that the intent, structure, and procedure of the commission was a farce, if not fraudulent, when considered from a scientific standpoint. If there is a failure here, it is the failure of the scientific community, especially behavioral scientists, to voice outrage over the manner in which basic scientific principles were disposed of by the commission, much as heretics at the Inquisition.

A necessary but not sufficient condition for a scientific investigation to occur is that the rules of basic logic be operable. Consider, then, the objectives of the commission as stated in the commission's charter (Department of Justice, 1985):

> The objectives of the Commission are to determine the nature, extent, and impact on society of pornography in the United States, and to make specific recommendations to the Attorney General concerning *more effective ways in which the spread of pornography could be contained,* consistent with constitutional guarantees. (p. 14, emphasis added)

The "nature, extent, and impact" are empirical questions open to scientific investigation. However, to recommend ways of containing the spread of pornography presupposes it needs to be contained, which in turn presupposes it is harmful. In other words, the question has been begged, and the requirement of logic is violated at the outset. It would have been much more honest for the commission to simply have stated its objective as finding ways of preventing the spread of pornography without violating First Amendment rights. There was never any intent to evaluate the scope and effect of pornography on society.

A second necessary but not sufficient condition for a scientific inquiry is that it be, to the extent possible, an objective inquiry. In this case, at the bare minimum, one would expect the members of the commission to be disinterested parties with an essentially neutral viewpoint. The composition of the Commission has been detailed elsewhere (e.g., Nobile & Nadler, 1986); suffice it to say there is no question that it was stacked with individuals who believed pornography was harmful to begin with, most of whom, indeed, having made this public knowledge prior to the formation of the commission. Judith Becker, a psychologist and the only commissioner with actual experience in dealing with sex offenders, was one of two dissenters from the commission's report.

In evaluating the scientific data on the effects of pornography, the commission was so stymied that they resorted to a new definition for evaluating the evidence, something the commission termed the "totality of evidence." This was a clear attempt to circumvent the lack of scientific evidence for harm by pornography and still find a causal relation between pornography and harm to society (Becker, 1986).

There are two other issues that I would raise with Page. First, he wrote, "This (upward) trend in the content of pornographic material is consistent with the Bureau of Justice's recent study, showing an increase in crime and violence generally in North America" (p. 579). In other words, there is a correlation. Somewhere here Page has forgotten that correlation may be a necessary but not sufficient condition for causation, but that it does not imply causation. The correlation between the incidence of rape in the United States and the membership in the Southern Baptist church is a highly significant .96, a figure easily calculated from rape statistics and membership data of that particular denomination. Would he contend that Southern Baptists should be outlawed because their numbers are associated with the increase in the incidence of rape? Of course not. The incidence of rape will correlate positively with any variable that also evidences an increase over the same time period.

Most important, Page did not consider the possibility that the primary reason researchers are more conservative (if not backpedaling) about their inferences in public testimony than in their publications of experimental results is because the original inferences went far beyond scientific credibility. I believe if Page critically examined the original research of the authors he takes to task in the same manner as he did the Linz, Donnerstein, and Penrod (1987) article he would understand, as I have demonstrated elsewhere (Mould, 1988a), that there are major difficulties with methodological, analytical, and interpretive aspects of almost all of this body of literature. A brief example is illustrative. Consider an underlying presupposition in the issue of depictions of cartoon violence in *Playboy* and *Penthouse:* Women are the victims and men are the perpetrators. It would seem common sense that this assumption would be valid. Malamuth and Spinner (1980) certainly assumed this in their highly publicized article on violence in these magazines. However, it is not true. Even a cursory perusal of the *Playboy* and *Penthouse* issues they studied will reveal that a substantial portion of cartoons depicting violence depict violence by women against men. An outstanding example of this is the cartoon heralding the advent of the Wicked Wanda comic strip in *Penthouse.* In the cartoon, Wicked Wanda (1987) is scantily clad and has two mastiffs on leashes chained to her belt. One of the dogs is at the throat of a man lying at the bottom of a short staircase, and has already shredded the man's clothing. The man's facial expression is one of terror, whereas Wicked Wanda's expression is one of bemused curiosity. The presence of images of women aggressing against men completely confounds any inferences of these images encouraging or facilitating sexual violence by men against women.

Much of the data regarding pornography has lain dormant over the last 15 years. It has surfaced as a consequence of the Meese commission as well as feminist critique. In its dormant form, overstatements, overgeneralizations, and glossing over of inconsistent and incongruent results served the useful purpose of stimulating further research and inspiring academic debate. Employing this data as a rationale for implementing public policy should frighten all concerned with free speech and individual liberty.

Instead of being berated, Linz et al. should be congratulated for retrenching their interpretations and bringing them more into line with their data.

REFERENCES

Becker, J. (1986, November). *The Presidential Commission on Pornography: Politics, pro-*