Donald Campbell and his colleagues have elaborated an approach to research based on the elimination of threats to four validity types: internal, external, construct, and statistical conclusion. The central thesis of this article is that the process of eliminating validity threats depends fundamentally on no-difference findings, a fact that, unfortunately, has not been made explicit by researchers. The implications of this neglect are explored using examples from a number of different substantive areas such as psychology, health, and medicine. Finally, the intrinsic role of no-difference findings is described in the context of all four validity types, and suggestions for improving the process are offered.

USE AND MISUSE OF NO-DIFFERENCE FINDINGS IN ELIMINATING THREATS TO VALIDITY

WILLIAM H. YEATON
University of Michigan, Center for Research on Utilization of Scientific Knowledge

LEE SECHREST
University of Arizona

A realistic characterization of a typical first reaction to the topic of "no-difference findings" might be something like: "Oh, yes, but that's not very important, and besides it's not something that I encounter very often, so I'll wait to try to understand it better."

This rationale is inaccurate on at least two counts. First, it is quite simple to construct a long list of important questions for which no-difference findings would be critical: Does Agent Orange increase the risk of health problems? Are the pollutants in Love Canal associated with an increased risk of genetic defects? Are children who are placed in

AUTHORS' NOTE: *This work was supported by grant HS 04825 from the National Center for Health Services Research. Thanks go to Mel Mark, David Ronis, and Amiram Vinokur for their comments on an earlier version of this article.*

EVALUATION REVIEW, Vol. 10 No. 6, December 1986 836-852

day-care centers as intellectually, socially, and emotionally advanced as children who remain in the home? Is a new, cheaper drug with fewer side effects as effective as the existing standard in preventing heart attacks? Does saccharin increase one's risk of developing bladder cancer? (Only the space requirements of this article limit the length of this list.)

The second way in which this rationale is inaccurate is a primary basis for the existence of this article. In fact, we argue that it is precisely because of the high frequency of occurrence that no-difference findings are so critical. What has long been ignored by researchers, however, is that the no-difference case is absolutely essential to establishing differences. In making the argument that a particular variable makes an important difference, one will routinely strive to minimize the importance of other variables. Said slightly differently, we maintain that in nearly all instances in which researchers wish to establish the difference case, they will attempt to show that there are no differences on other confounding or "extraneous" factors.

This approach, as described briefly, can be traced directly to Donald Campbell and his colleagues. Before addressing the fundamental intertwining of the Campbellian approach and no-difference logic, however, we wish to digress to discuss some of its important elements.

Central to Campbell's approach from the very beginning has been the concept of validity. In what is generally acknowledged to be the first relatively complete statement of his approach (Campbell and Stanley, 1963), special emphasis was placed on the importance of validity in assessing the quality of research. Two kinds of validity, internal and external, were introduced in the now classic *Experimental and Quasi-Experimental Designs for Research*, which Campbell coauthored with Julian Stanley. In subsequent writings (Cook and Campbell, 1979), two other kinds of validity, statistical conclusion and construct, were introduced. These four kinds of validity took on particular importance as they structured much of Campbell's writings as well as the vast literature describing and elaborating the approach.

The task of research from a Campbellian perspective is to establish each of the four kinds of validity. To accomplish this task, research should aim to eliminate or at least render less plausible specific threats to each of the validity types. These threats or "plausible rival hypotheses" have to be warded off to make a convincing argument to the reader that the author's interpretation is the most believable one.

This article focuses on the different methods that researchers have used to eliminate these threats to validity and centers attention on the

intrinsic role of no-difference findings in these methods. We argue that when researchers try to eliminate validity threats, their usual strategy is to produce a no-difference finding and to use this no-difference result as the crux of their argument that the validity threat is unlikely to have accounted for a particular set of results.

The argument presented to make the no-difference case is either informal, as when authors state that patients in two groups appear to be similar, or formal, as when a test of statistical significance is conducted. Unfortunately, both arguments suffer inherent weaknesses. In the first instance, one cannot be certain that the apparently similar groups are similar enough or that the groups only appear similar to the researchers conducting this study. In the second case we cannot be certain that the groups are equivalent since, as statisticians are quick to point out, we can never technically prove the null hypothesis.

Although the elimination of validity threats is routinely accepted by most social scientists, it unfortunately rests on a very shaky foundation, namely reliance on no-difference findings. In contrast to the time-tested and relatively uncontroversial procedures enacted to demonstrate differences, there is no general agreement regarding the most appropriate methods to utilize when the researcher's aim is to show that no difference exists (Blackwelder, 1981).

We do not wish to be seen as arguing that the goal of producing no-difference findings is somehow incorrect, but we do maintain that when compared to the rather well-established method of conducting research whose purpose is to show a difference, inferential problems of no-difference results are not well understood. Part of this deficiency rests with the fact that most researchers concentrate on study outcomes meant to show a difference. We argue elsewhere (Sechrest and Yeaton, 1982) that the subset of studies whose intent is to produce no-difference outcomes is itself important but maintain that the lack of attention to no-difference research only exacerbates the problems with proper use of no-difference findings in eliminating validity threats.

The general purpose of this article is to identify some of the ways that no-difference results are used to eliminate validity threats in social, health, and medical research. To our knowledge, there have been no previous articles that directly acknowledge the importance of no-difference findings in eliminating validity threats. We also discuss various problems with this approach, point out implications of incorrect application of the strategy, and offer several suggestions for improvement.

VALIDITY THREATS AND THE USE OF NO-DIFFERENCE FINDINGS

INTERNAL VALIDITY

Internal validity is concerned with the causal relationship between two variables. To the extent that internal validity is high, other factors can be ruled out as causing the outcomes of interest. Thus, the independent variable and not some other "extraneous" or unspecified variable is said to be causal.

In many research studies, the single most important threat to internal validity is *selection*. Selection bias occurs when the participants in two groups differ in some aspect that could explain outcome differences between groups. Strategies such as randomization and matching are designed to produce groups that are equivalent. Thus, where possible, research that compares outcomes between two or more groups will attempt to show that there are no initial differences between groups. In other instances (for example, the nonequivalent control group design), researchers will try to make the case, perhaps by statistical adjustments, that any preexisting differences would not likely account for posttest differences.

There is a wide range of techniques whose purpose is to demonstrate that groups are equivalent. In the prototypical two-group medical study, numerous baseline measures are monitored and tested for statistical significance. For example, in a randomized, double-bind trial of nifedipine in unstable angina (Gerstenblith et al., 1982), there were no initial, statistical differences between the nifedipine group and a placebo group in age, sex, race, duration of angina, prevalence of diabetes, or in 13 other relevant patient characteristics. By contrast, baseline comparability of 19 different characteristics of patients who were referred either to a systematic antihypertensive treatment program or to community medical therapy (Hypertension Detection and Follow-up Program Cooperative Group, 1979) was not established with tests of statistical significance but rather with the statement, "Differences between the two groups for these and other variables were small, indicating the effectiveness of the randomization procedure." Similarly, Doherty and Dieppe (1981) did not test for any initial differences between the left and right knee of patients with arthritis but randomly assigned each knee to receive either yttrium-90 plus steroid or a saline

solution plus steroid. The authors apparently assumed that there was no important difference between the knees of each individual patient.

In those instances for which persons have not been randomly assigned to groups, selection biases are increasingly likely, making it even less certain that initial group differences are ruled out as a possible explanation of results. Again, no-difference findings play a vital role in making the case that selection is not a threat to internal validity. In studies of the relationship between aspirin intake and Reye's syndrome, researchers go to great lengths to attempt to match children who have contracted Reye's syndrome with control patients who are judged to be alike in all other ways except for not contracting the disease (for instance, see Waldman et al., 1982). Unfortunately, matching does not guarantee group equivalence, especially when there are many variables to match on or when one may have failed to match groups on a critical but unknown variable.

The rationale for using historical control groups rests on the same desire to avoid selection bias since a much more convincing case can be argued with the inclusion of an equivalent comparison group. Unfortunately, historical controls very often distort outcome differences between groups in controlled studies despite the apparent absence of initial differences between treatment groups and historical controls (Sacks et al., 1982).

Most applied researchers are well aware of the fact that the process of randomization will not necessarily eliminate differences between groups. Small sample size or chance alone may be the culprit in producing differences. Chalmers and his colleagues illustrate a more subtle way that difference may emerge despite randomization (Chalmers et al., 1983). Chalmers et al. found that at least one prognostic variable was maldistributed between groups in only 14.0% of a sample of 102 randomized trials of myocardial infarction if the investigators were blinded with respect to treatment assignment at the time the patient was to be included in the study. In contrast, the percentage was 26.7 in randomized but unblinded studies. Thus, the no-difference case may not be established, and selection bias can still occur in the context of randomization if patient recruitment has not proceeded in a blinded fashion. Without blinding, physicians may subtly encourage or discourage patients with particular profiles to enter experimental or control groups.

Another common threat to internal validity is termed *maturation*. In many educational interventions it is important to establish that students

in both experimental and control groups have similar developmental histories with regard to cognitive ability. This substantiates the fact not only that they have equal abilities but also that they have equal rates of change in ability levels. Similarly, in medical research it may be important to show that patient groups have similar illness trends and that one group is not developing unhealthy symptoms faster than the other. Dikmen et al. (1983) make a similar point in emphasizing the importance of testing rates of improvement in patients suffering from head injury. In both of these instances trend tests should be utilized to show that there are no differences between groups.

Research with protracted follow-up periods may be susceptible to the threat of *mortality*, or as it is often called, attrition. From an internal validity point of view, when patients with differing characteristics are lost from the groups of a study (differential attrition), the original equality of groups is compromised, thus allowing critics to argue that group differences at follow-up could have accounted for outcome differences. In a randomized controlled study of the ability of aspirin to prevent deaths in patients who had suffered a previous myocardial infarction, there were numerous differences in the reasons that patients from the aspirin groups and those from the comparison group receiving oral anticoagulants had withdrawn from the study (EPSIM Research Group, 1982). However, none of these differences were tested for statistical significance.

Differential attrition is an especially thorny problem in controlled studies of coronary artery bypass graft surgery since long-term follow-ups are necessary to evaluate surgery. The problem is still more complicated, however, because those patients who have not received surgery but instead have remained on a drug regimen not only drop out of the medical group but also cross over to become members of the surgical group, further distorting the possible nonequivalence between groups. Since it may be those medical patients who have the most severe angina who become crossovers (Murphy et al., 1977), it becomes still more difficult to make unambiguous claims for the effectiveness of surgery. Unfortunately, although the problem of crossovers is acknowledged and the number of crossovers typically reported (see, for example, European Coronary Surgery Study Group, 1982), data describing the characteristics of crossovers are typically lacking. Without these data it is impossible to establish that there are no statistically significant differences between groups in the pertinent characteristics of those who are lost to the study.

Problems with the use of no-difference findings to eliminate threats to internal validity. The strongest case can be made for the legitimacy of no-difference findings to rule out threats to internal validity when appropriate comparisons are an integral part of the research methodology rather than an afterthought. Post hoc comparisons of the respective composition of groups are simply not as convincing as planned comparisons of possible group differences. Yet some studies may be retrospective and, thus, necessarily entail intact groups. This research does not allow a priori equalization but instead requires creative statistical analysis to render selection less plausible. Of course, if the specific group differences are not possible causes of differences in outcomes, then statistical adjustment is not necessary. It is not our intent to diminish the value of these "patched-up" designs but rather to urge that they be conducted only when there is no alternative.

In those instances in which a statistical test is used to argue for the equivalence between comparison groups, care must be taken to ensure that adequate statistical power is available. Quite often one will attempt to show that comparison groups do not differ with regard to sex, race, IQ, or other relevant demographic variables. In addition, it is important that statistical tests between subgroups of comparison groups have sufficient sample size to detect initial group differences. Unfortunately, researchers appear to be much more careful to ensure that between-group statistical tests of outcome variables have sufficient power than to make certain that tests of initial differences between subgroups on demographic or clinical variables have sufficient power. Given that subgroups will necessarily have smaller Ns than entire groups, the deck is stacked to show no initial differences between subgroups.

Upon reflection, we cannot recall a single instance in which authors have lamented the lower power available in tests of initial differences to rule out threats to internal validity such as selection. Furthermore, patient heterogeneity will necessarily inflate error terms and make it more likely that differences are not significant. On the other hand, lack of statistical power due to patient heterogeneity or small sample size is often used as an excuse when difference outcomes are sought.

From a different perspective, since many group characteristics are typically tested for initial differences (such as age, sex, aspects of medical history), it would be expected that a few spurious differences would be produced by chance alone. *If* such subgroups have adequate sample size and statistical power, then it is legitimate to cite the problem of experiment-wise-error rate and to opt for such standard approaches as lowering the alpha level of each comparison.

STATISTICAL CONCLUSION VALIDITY

Statistical conclusion validity refers to the legitimacy of inferences based on statistical tests of significance. Statistical conclusion validity is particularly important in the case of no-difference findings since there are so many factors that facilitate incorrect, no-difference conclusions (such as low statistical power). This section of the article first focuses on the familiar case in which study outcomes are of the no-difference variety. Later, a more neglected aspect is emphasized: the fundamental role of no-difference findings in actually establishing statistical conclusion validity.

Studies yielding no difference assume many forms. For example, much of the research comparing rival treatments for breast cancer show that radical mastectomy and other invasive forms of treatment do not improve survival benefit beyond that produced by less invasive surgery with and without radiation (Fisher et al., 1985). In a randomized, double-blind study to test the effectiveness of acupuncture on osteoarthritic pain (Gaw et al., 1975), responses to acupuncture and to a control, sham treatment were shown to be equivalent. Sometimes the aim is to compare the rates of two or more curves, as was the case in a study of female fecundity as a function of age (Federation CECOS et al., 1982). Results indicated that cohorts 25 years of age and younger and cohorts 26 to 30 years of age did not differ statistically in their rate of pregnancy whereas older cohorts had a significantly reduced likelihood of pregnancy. Finally, researchers may be interested in showing the similarity between personality profiles (see Cronbach and Gleser, 1953) such as the MMPI profiles of murderers and rapists.

Before tests of outcome differences are conducted, however, researchers must show that the assumptions of tests of significance are satisfied. This is where the critical role of no-difference findings in establishing statistical conclusion validity comes into play. For example, analysis of variance requires that the variance in different groups be homogeneous—that there be no significant differences between the variances in each group (Scheffé, 1959). Similarly, the independence assumption must be satisfied. In the context of this article, one could translate this to mean that a nonsignificant amount of dependence must be present for a legitimate comparison to be made. The independence assumption is especially problematic in medical research since results from the same patient may be reported in more than one study. One of the authors (Wortman and Yeaton, 1983) confronted this problem in a synthesis of

the outcomes of controlled trials of coronary artery bypass surgery. One way of demonstrating that the independence assumption is satisfied might be to show that the percentage of duplicate patients in the studies being compared is not significantly different from zero.

On occasion, there may exist some uncertainty about which of several statistical analyses is most appropriate to a given data set. In such an instance one viable strategy is to conduct several analyses and determine if the different analytic strategies yield similar results, or at least the same pattern of results. Wortman and Yeaton (1983) used this strategy in evaluating results of controlled trials of coronary artery bypass-graft surgery, finding a consistent survival advantage to surgery as compared to a drug regimen. In the face of such consistency (no-difference in outcomes using different analyses), inferential power is considerably bolstered and statistical conclusion validity is enhanced.

Problems with the use of no-difference findings in the context of statistical conclusion validity. Certainly, issues such as statistical power are particularly important with regard to tests of no-difference between groups. Technically, statistical power is a function of alpha, effect size, and sample size (Cohen, 1977) as well as other methodological criteria (Sechrest and Yeaton, 1982) including measurement sensitivity (Lipsey, 1983). One simply cannot assess the adequacy of no-difference findings without first checking these factors.

One potential problem with the use of no-difference findings relates to the common practice of pooling nonsignificant results before conducting general statistical tests (Bancroft and Han, 1983). In the example described previously in which tests of initial differences were made between various group strata, absence of difference means that subgroup scores can be pooled and overall group results can be tested. Used appropriately, pooling raises degrees of freedom and statistical power. However, it is not at all clear what steps should be followed if several tests are "close" to statistical significance or what proportion of significant tests between subgroups precludes the practice of pooling.

CONSTRUCT VALIDITY

Construct validity is concerned with the mechanism by which change manifests itself. It seeks to establish the reasons for change rather than directly probe the existence of causal inferences or the degree to which appropriate statistical issues have been addressed.

Researchers use a number of no-difference strategies to enhance the construct validity of research findings. In studies of the effectiveness of psychotherapy, contact control groups are used to equate the amount of attention received by treatment and comparison groups (see Beck et al., 1984). In this way, the specific elements that characterize psychotherapy rather than the degree of attention given to patients can be argued to produce clinically important change. Thus, to establish construct validity researchers will attempt to show that there is no difference between groups on any variables other than the independent variables that may be construed as causal.

Precisely the same rationale is used in medical research when placebo or sham treatment groups are used to minimize the plausible role of extraneous treatment dimensions. Carpenter and his colleagues (Carpenter et al., 1983) provided a stern test of the claim for beneficial effects of hemodialysis with schizophrenics by using a sham control group. In this case as well as the psychotherapy example cited earlier, treatment equivalence was built into the research design. In most instances specific tests of no difference are not conducted (for example, in the amount of attention given or the number of relevant components removed from the blood), though they certainly could and should be.

Construct validity can be enhanced considerably by appropriate use of no-difference findings. In their efforts to document the effectiveness of the British Road Safety Act of 1967, Ross et al. (1970) were able to show that the number of traffic fatalities decreased during evening hours when pubs were open but did not decrease during commuting hours. The same strategy was employed by Parker (1963), who demonstrated that the introduction of television into the community decreased the number of fiction books checked out of the local library without a concurrent change in the number of non-fiction books read.

More generally, Cook and Campbell (1979) have argued that construct validity might be assessed by two processes: "first, testing for a convergence across different measures or manipulations of the same 'thing', and second, testing for a divergence between measures and manipulations of related but conceptually different 'things.' " When one realizes that convergence is synonymous with no difference, the integral role of no-difference findings in this assessment process is blatantly obvious.

In many clinical trials, maintenance of double-blind conditions (neither the provider nor recipient of treatment is aware of the treatment that he or she receives) is necessary to maintain construct validity. To

test for maintenance of double-blindness, participants in all study groups are asked to identify their treatment group assignment (for instance, active drug or placebo). The lack of a significant difference between study groups in these rates of identification suggest that expectation effects have been equalized. Disparity in these rates of identification by either patients or physicians (see Byington et al., 1985) suggest that double-blindness has not been maintained and that construct validity has been weakened.

Construct validity may be called into question when the independent variable of interest or a conceptually related independent variable "contaminates" the comparison group. For example, in the case-control study of cytogenetic findings of persons living near Love Canal and a matched group geographically removed from Love Canal (Heath et al., 1984), it was important to establish that there had been no inequity in exposure to radiation or other toxic chemicals that might be associated with chromosomic aberrations. Based on questionnaire evidence, the authors reported no significant differences in extraneous exposure. In the previously mentioned study of the effects of aspirin on patients with prior myocardial infarctions (EPSIM, 1982), the researchers reported that "the use of drugs that might have affected survival, such as beta-blocking agents or antidysrhythmics, was fairly well balanced in the two groups" and mortality results were statistically adjusted for the use of concomitant medications. To demonstrate the blinding of the placebo condition had been maintained and, therefore, that subtle expectations could not account for differences, researchers in the Lipids Research Clinics Program (1984) compared the percentage of clinic staff and participants in the drug and placebo groups that could correctly identify their respective treatment assignments. Although the differences appeared small (less than 3%), they were not tested for significance.

Problems with the use of no-difference findings in the context of construct validity. When two or more medical treatments are compared, it is critical to ensure that treatment protocols are adhered to since departures from protocol will compromise the construct validity of the findings. One prominent difficulty in maintaining treatment adherence, or "integrity" as we have termed it in our previous writings (Yeaton and Sechrest, 1981), is that it may not be equally easy to maintain integrity for all treatments. To illustrate, it is almost certainly easier to achieve higher levels of treatment integrity for drug treatments of hyperactivity than for behavioral interventions. Unless there is no substantial

difference between the level of integrity for each treatment, it will be impossible to reach an unambiguous conclusion regarding relative effectiveness.

EXTERNAL VALIDITY

External validity is concerned with the generalizability of research findings, and thus asks whether similar results would apply to persons, situations, and time periods other than those examined in the study. In the context of no-difference results the task is straightforward—establish that the persons, situations, and time periods to which one wants to generalize are similar enough to those chosen to convince the reader that external validity will be high.

Probably the most common method of establishing comparability (no difference) between population and sample units (persons, responses, situations, and time periods) is to choose randomly from the relevant population. The primary advantage of random sampling lies in its ability to produce samples of persons with characteristics closely comparable to those of given populations. For example, by randomly choosing respondents to health-care surveys, researchers maintain that findings apply to larger units of interest (such as cities, states, or countries). Even when random sampling is not utilized, researchers will often try to establish in an informal way that there is no difference between persons in the experiment and persons in the extraexperimental world to which one wants to generalize. In this case researchers may state the demographic dimensions along which the two groups are comparable (racial mix, age, SES, and so on).

As in each of the earlier-described threats to validity, researchers use the logic of no-difference findings to establish external validity. For example, in those instances for which there have been losses to follow-up, researchers will try to show that the proportions are essentially equal in the group of dropouts and the group available at follow-up. (We assume that the characteristics of those who leave the study are the same in both groups so that internal validity is not compromised. Here, the relative incidence of losses between groups is at issue.) For example, Beard et al. (1982) compared the effects of a no-added-sodium diet in patients with mild hypertension to results found in control patients, noting that the overall rate of attrition due to exclusions and dropouts was similar in the two groups (19.6% and 21.1%, respectively) but did not provide a statistical test of the difference.

No-difference logic may also be used to test the impact of losses of patients before assignment to comparison groups. Charlson and Horowitz (1984) studied trials listed in the 1979 inventory of the National Institutes of Health and found that only a few trials documented the characteristics of patients who were eligible but not entered into the trials. Without demonstrating the essential equivalence between included and excluded participants on variables related to outcome, it will be impossible to make a persuasive case that results will generalize to a relevant population.

The strength of generalizations can be enhanced if both the pattern of findings and patient characteristics are comparable in the sample and the population. For example, in the case of coronary heart disease patients, if the survival curve of patients who have received a drug—perhaps one of the beta-blockers such as propranolol—closely approximates the survival curve of similar patients in a normative comparison group who do not suffer from coronary artery heart disease, then benefits of this specific magnitude may be generalized to other coronary heart disease patients with similar symptomatology. Without matching both patients and the pattern of results, such inferences are more likely to be called into question.

One of the advantages of multicenter clinical trials is the potential of generalizing results to other centers with no substantial difference between facilities or physician competence. In the CASS study (1983) of coronary artery bypass surgery with patients having mild coronary heart disease, data were taken from 15 separate locations representing a range of medical centers and physicians, thus allowing broad generalization of findings.

Problems with the use of no-difference findings in the context of external validity. The problems that arise in establishing external validity closely parallel those encountered when one wishes to argue against selection as a plausible rival hypothesis in establishing internal validity. In both cases researchers will attempt to show that relevant groups do not differ along dimensions that may contribute to the outcomes found. For example, just as adequate statistical power is required in statistical tests of no difference to establish internal validity, so too will it be necessary to achieve high levels of power to establish external validity.

IMPLICATIONS AND CAVEATS

We would hope that the preceding discussion makes explicit the fundamental and pervasive role of no-difference findings in eliminating a variety of validity threats. This knowledge should be of value to researchers who wish to plan a convincing study as well as to the consumer of research findings who wants to assess the believability of research. For the researcher the goal will be to design a study to maximize the possibility of valid difference or no-difference findings and thus to minimize validity threats. For the consumer the goal will be to evaluate the findings presented in the study as evidence and to judge whether particular validity threats are unimportant. In both cases no-difference findings are intrinsic to researcher and consumer goals.

Nowhere in the discussion have we meant to imply that a validity threat is eliminated if it can be shown that a statistical test reveals no difference. Although this might be regarded as a necessary condition in many cases, we do not wish to be seen as arguing that statistical insignificance is sufficient evidence to assume that a validity threat has been eliminated. Certainly, consistency of no-difference results with previous research should enhance one's argument considerably. For example, the absence of sex differences in previous studies of suggestibility (Maccoby and Jacklin, 1974) offers a convincing context for accepting the lack of sex differences within a particular study. Current theory concerning the etiology of breast cancer emphasizes its systemic nature, making no-difference findings between rival treatments such as radical mastectomy and lumpectomy (Fisher et al., 1985) much more believable since both of these treatments are targeted toward a local rather than a systemic view of the disease.

The relationship between no-difference findings and the elimination of validity threats has considerable implication for one's general perspective of scientific research. Since we subscribe to the Campbellian notion that good research renders rival hypotheses less plausible, we do not hold to a plausible-implausible dichotomy, but instead prefer to think in terms of a continuum. From the gradual accretion of evidence from multiple studies, each with complementary weakness, alternative explanations become less plausible by degree and the veridicality of each relationship comes to be accepted (Staines, 1974). Since the reader may be inclined to form some sort of cognitive equivalence between plausible-implausible and difference-no difference, there is a real danger

that viewing difference-no difference as a dichotomy will reinforce a similar dichotomous view with respect to plausibility. We do not wish to perpetuate what we believe to be a mistaken belief.

Finally, we hope that our discussion of the uses of no-difference findings will serve as an impetus to other researchers. Certainly, there is much room for improvement in developing a more coherent though still diversified strategy for making the no-difference case. Yet we are optimistic that this strategy will improve given the fundamental intertwining of the threats to validity approach and the role of no-difference findings. Such a relationship seems to guarantee that neither can be improved without an accompanying benefit to the other.

REFERENCES

BANCROFT, T. A. and C. P. HAN (1983) "A note on pooling variances." J. of the Amer. Statistical Assn. 78: 981-983.

BEARD, T. C., H. M. COOKE, W. R. GRAY, and R. BARGE (1982) "Randomized controlled trial of a no-added-sodium diet for mild hypertension." Lancet ii: 455-458.

BECK, J. G., F. ANDRASIK, and J. G. ARENA (1984) "Group comparison designs," in A. S. Bellack and M. Hersen (eds.) Research Methods in Clinical Psychology. New York: Pergamon.

BLACKWELDER, W. C. (1981) "'Proving the null hypothesis' in clinical trials." Controlled Clinical Trials 3: 345-353.

BYINGTON, R. P., J. D. CURB, and M. E. MATTSON (1985) "Assessment of double-blindness at the conclusion of the beta-blocker heart attack trial." J. of the Amer. Medical Assn. 253: 1733-1736.

CAMPBELL, D. T. and J. C. STANLEY (1963) Experimental and Quasi-Experimental Designs for Research. Chicago: Rand McNally.

CARPENTER, W. T., J. H. SADLER, P. D. LIGHT, T. E. HANLON, A. A. KURLAND, M. W. PENNA, W. P. REED, E. H. WILKINSON, and J. J. BARTKO (1983) "The therapeutic efficacy of hemodialysis in schizophrenia." New England J. of Medicine 308: 669-675.

CASS Principal Investigators and their associates (1983) "Coronary Artery Surgery Study: a randomized trial of coronary artery bypass surgery. Quality of life in patients randomly assigned to treatment groups." Circulation 68: 951-960.

CHALMERS, T. C., P. CELANO, H. S. SACKS, and H. SMITH (1983) "Bias in treatment assignment in controlled clinical trials." New England J. of Medicine 309: 1358-1361.

CHARLSON, M. E. and R. I. HOROWITZ (1984) "Applying results of randomized trials to clinical practice: impact of losses before randomization." British Medical J. 289: 1281-1284.

COHEN, J. (1977) Statistical Power Analysis for the Behavioral Sciences. New York: Academic Press.

COOK, T. D. and D. T. CAMPBELL (1979) Quasi-Experimentation: Design and Analysis Issues for Field Settings. Chicago: Rand McNally.

CRONBACH, L. J. and G. C. GLESER (1953) "Asessing similarity between profiles." Psych. Bull. 50: 456-473.

DIKMEN, S., R. M. REITAN, and N. R. TEMKIN (1983) "Neuropsychological recovery in head injury." Archives of Neurology 40: 333-338.

DOHERTY, M. and P. A. DIEPPE (1981) "Effect of intra-articular yttrium-90 on chronic pyrophosphate arthropathy of the knee." Lancet ii: 1243-1246.

E.P.S.I.M. Research Group (1982) "A controlled comparison of aspirin and oral anticoagulants in prevention of death after myocardial infarction." New England J. of Medicine 307: 701-708.

European Coronary Artery Surgery Study Group (1982) "Long-term results of prospective randomized study of coronary artery bypass surgery in stable angina pectoris." Lancet ii: 1173-1180.

Federation CECOS, D. SCHWARTZ, and M. J. MAYAUX (1982) "Female fecundity as a function of age. Results of artificial insemination in 2193 nulliparous women with azoospermic husbands." New England J. of Medicine 306: 404-406.

FISHER, B., M. BAUER, R. MARGOLESE, R. POISSON, Y. PILCH, C. REDMOND, E. FISHER, N. WOLMARK, M. DEUTSCH, E. MONTAGUE, E. SAFFER, L. WICKERHAM, H. LERNER, A. GLASS, H. SHIBATA, P. DECKERS, A. KETCHAM, R. DISHI, and I. RUSSELL (1985) "Five-year results of a randomized clinical trial comparing total mastectomy and segmental mastectomy with or without radiation in the treatment of breast cancer." New England J. of Medicine 312: 665-673.

FISHER, B., C. REDMOND, E. R. FISHER, M. BAUER, N. WOLMARK, D. L. WICKERHAM, M. DEUTSCH, E. MONTAGUE, R. MARGOLESE, and R. FOSTER (1985) "Ten-year results of a randomized clinical trial comparing radical mastectomy and total mastectomy with or without radiation." New England J. of Medicine 312: 674-681.

GAW, A. C., W. C. LENNIG, L. W. CHANG, and L. C. SHAW (1975) "Efficacy of acupuncture on osteoarthritic pain. A controlled, double-blind study." New England J. of Medicine 293: 375-378.

GERSTENBLITH, G. et al. (1982) "Nifedipine in unstable angina. A double-blind, randomized study." New England J. of Medicine 306: 885-889.

HEATH, C. W., M. R. NADEL, M. M. ZACH, A.T.L. CHEN, M. A. BENDER, and J. PRESTON (1984) "Cytogenetic findings in persons living near the Love Canal." J. of the Amer. Medical Assn. 251: 1437-1440.

Hypertension Detection and Follow-up Program Cooperative Group (1979) "Five-year findings of the Hypertension Detection and Follow-up Program. I. Reduction in mortality of persons with high blood pressure, including mild hypertension." J. of the Amer. Medical Assn. 242: 2562-2571.

Lipid Research Clinics Program (1984) "The Lipid Research Clinics coronary primary prevention trial results. Reduction in incidence of coronary heart disease." J. of the Amer. Medical Assn. 251: 351-364.

LIPSEY, M. W. (1983) "A scheme for assessing measurement sensitivity in program evaluation and other applied research." Psych. Bull. 94: 152-165.

MACCOBY, E. E. and C. N. JACKLIN (1974) The Psychology of Sex Differences. Stanford, CA: Stanford Univ. Press.

MURPHY, M. L., H. N. HULTGREN, K. DETRE, J. THOMSEN, and T. TAKARO (1977) "Special correspondence: a debate on coronary bypass." New England J. of Medicine 297: 1470.

PARKER, E. B. (1963) "The effects of television on public library circulation." Public Opinion Q. 27: 578-589.

ROSS, H. L., D. T. CAMPBELL, and G. V GLASS (1970) "Determining the social effects of a legal reform: the British 'breathalyzer' crackdown of 1967." Amer. Behavioral Scientist 13: 493-509.

SACKS, H., T. C. CHALMERS, and H. SMITH (1982) "Randomized versus historical controls for clinical trials." Amer. J. of Medicine 72: 233-240.

SCHEFFÉ, H. (1959) The Analysis of Variance. New York: John Wiley.

SECHREST, L. and W. H. YEATON (1982) "Magnitudes of experimental effects in social science research." Evaluation Rev. 6: 579-600.

———(forthcoming) "Role of no-difference findings in medical research."

STAINES, G. L. (1974) "The strategic combination argument," pp. 417-430 in W. Leinfellner and E. Kohler (eds.) Developments in the Methods of Social Sciences. Dordrecht, Holland: D. Reidel.

WALDMAN, R. J., W. N. HALL, H. McGEE, and G. VAN AMBURG (1982) "Aspirin as a risk factor in Reye's syndrome." J. of the Amer. Medical Assn. 247: 3089-3094.

WORTMAN, P. M. and W. H. YEATON (1983) "Synthesis of results in controlled trials of coronary artery bypass graft surgery," pp. 536-551 in R. J. Light (ed.) Evaluation Studies Review Annual, Vol. 8. Beverly Hills, CA: Sage.

YEATON, W. H. and L. SECHREST (1981) "Critical dimensions in the choice and maintenance of successful treatments: strength, integrity and effectiveness." J. of Consulting and Clinical Psychology 49: 156-167.

William H. Yeaton is Assistant Research Scientist at the Institute for Social Research at the University of Michigan. His current research interests include medical technology assessment and evaluation research methodology, especially in the area of health.

Lee Sechrest is Chair of the Department of Psychology at the University of Arizona. His current research interests include the relationship between the quality of research methods and research outcomes and the utilization of these findings with regard to policy.