PERSONNEL PSYCHOLOGY
1996, 49

# THE PERCEPTIONS AND USAGE OF STATISTICAL POWER IN APPLIED PSYCHOLOGY AND MANAGEMENT RESEARCH

MARK A. MONE
School of Business Administration
University of Wisconsin-Milwaukee

GEORGE C. MUELLER
School of Business Administration
University of Wisconsin-Milwaukee

WADE MAULAND
School of Education
University of Wisconsin-Milwaukee

We first assess the current level of statistical power across articles in seven leading journals that represent a broad sample of applied psychology and management research. We next survey the authors of these articles to examine their perceptions and usage of statistical power analysis. Finally, we examine the perceptions and usage of power analysis in a survey of authors of regression-based research appearing in leading journals. Findings from the assessment of power and surveys of researchers indicate that power analyses are not typically conducted, researchers perceive little need for statistical power, and power in published research is low. We conclude by discussing implications of low power for the field and recommending avenues for improving researchers' awareness and usage of statistical power.

The predominance of a conservative approach to hypothesis testing, which places greater emphasis on reducing Type I errors than Type II errors, has been widely recognized in social science research (e.g., Chase & Chase, 1976; Cohen, 1988; Lykken, 1968; Rossi, 1990; Sauley & Bedeian, 1989). Whereas Type I errors can be considered *mistakenly believing* there are significant differences when they actually do not exist, Type II errors may be viewed as *mistakenly overlooking* significant differences that do exist. Given our field's emphases on minimizing Type I ($\alpha$) errors, some consequences of neglecting Type II ($\beta$) errors might be

The authors gratefully acknowledge comments on earlier drafts of this paper by Douglas Baker, Vincent Barker, Tom Becker, Tailan Chi, Michael Hitt, William McKinley, Paul Nystrom and Masoud Yasai. Contributions by Donna Randall were particularly helpful during the initial stages of this research.

Correspondence and requests for reprints should be addressed to Mark A. Mone, P.O. Box 742, School of Business Administration, University of Wisconsin-Milwaukee, Milwaukee WI 53201.

useful to acknowledge. One potential consequence is insufficient statistical power, usually due to small sample sizes. When statistical power is inadequate, researchers are unable to detect or reject actually false null hypotheses, meaning that they have overlooked meaningful differences or effects. As a result, Cohen (1962) has argued that much research reports spuriously "negative" results and there may frequently be "premature abandonment of useful lines of investigation" (p. 153). Another potential consequence of neglecting Type II errors, arguably less common, is excessively large data collections. Overpowered samples can be wasteful in terms of unnecessary effort, time, and resources spent, and may be oversensitive to trivial or irrelevant "significant" findings.

Despite the importance of statistical power, researchers routinely fail to plan studies with sufficient power. Over 30 years ago, Cohen (1962) noted that investigators contributing to the *Journal of Abnormal and Social Psychology* had, "on the average, a relatively (or even absolutely) poor chance of rejecting their major null hypotheses, unless the effect they sought was large" (p. 151). A recent update of this line of work suggests that the situation has not improved and may have, in fact, worsened (Sedlmeier & Gigerenzer, 1989). Insufficient statistical power has been the common finding in power assessments across diverse fields such as marketing (Sawyer & Ball, 1981), applied psychology (Chase & Chase, 1976), clinical psychology (Rossi, 1990), and communication (Chase & Tucker, 1975). As in other social science and behavioral fields, contemporary research in applied psychology and management rarely refers to the issue of statistical power. For example, the most recent published empirical assessments of statistical power in management research were conducted by Mazen, Graf, Kellogg, and Hemmasi (1987) and Mazen, Hemmasi, and Lewis (1987). They found that small effect sizes were the norm, small and medium effect size studies were characterized by limited statistical power, and explicit consideration of power issues was almost nonexistent.

Due to the importance of statistical power in applied psychology and management research and the limited research attention the topic has received, we think it is useful to further explore statistical power. Specifically, one purpose of the present study is to provide a current, representative assessment of the field's statistical power. To expand the scope of this study, we define management broadly to include organizational behavior, organizational theory, applied (I/O) psychology, human resource management, and strategic management. Most research on statistical power in these areas has been based on methodological or computational aspects (e.g., Cascio & Zedeck, 1983; Stone-Romero & Anderson, 1994), rather than assessments of power levels. Further, recent assessments of statistical power in management journals are now dated, cover a limited domain, and may not be widely generalizable. Mazen et al. (1987b), for example, examined data from 3 years (1982–1984) of *Strategic Management Journal* and one year (1984) of *Academy of Management Journal*, whereas Mazen et al. (1987a) examined data published in 1984 by the *Academy of Management Journal*, the *Journal of Management* and the *Proceedings of the Midwest Division of the Academy of Management*. In short, we do not know whether past examinations of statistical power in management research are representative of current practices.

The second purpose of our study is to explore researchers' perceptions and usage of statistical power analysis. Our goal is to go beyond the assessment of power levels in published research by investigating underlying beliefs that lead to specific research behaviors. Such an investigation may be a better source for understanding the *extent of power analysis usage*, relative to actual power assessments, and may also enable more accurate judgments concerning the impact of past calls for greater statistical power. Finally, an examination of perceived purposes and importance of statistical power may allow greater *discernment of reasons leading toward or away from statistical power analysis* in planning or evaluation stages of research. For example, are researchers aware of power, but consciously choosing not to conduct power analyses? If so, why? Answers to such questions represent an opportunity to understand and assess the accuracy of perceptions concerning the role and use of statistical power analyses.

A third purpose of this research is to call attention to the role of statistical power in terms of how researchers approach it and, more broadly, its implications for the field of management research. If researchers deem statistical power to be unimportant and subsequently choose to ignore it, what are the consequences for the field? The costs and implications for knowledge advancement of such a situation will be explored, and recommendations will be offered with respect to improving both awareness and usage of statistical power.

In the following section, we first describe an assessment of current statistical power levels in leading management journals. We then report the findings of two surveys of authors publishing empirical research in leading journals. The first survey focuses on perceptions and usage of leading researchers using a variety of empirical techniques; the second focuses on authors employing management's dominant analytic technique, multiple regression/correlation analyses. Our study draws from three independent sources to triangulate different perspectives on the same phenomenon, following examples by Orr, Sackett, and Dubois (1991) and Weinzimmer, Mone, and Alwan (1994) and suggestions by proponents of multiple methods and replication studies (Jick, 1979; Mone & McKinley, 1993; Neuliep, 1991; Rosenthal, 1990).

*Examination of Power Levels, Usage, and Perceptions*

*Study 1: Statistical Power in Management Journals*

*Article selection.* To assess the current level of statistical power in management research, we examined empirical articles in seven leading management journals, including *Academy of Management Journal, Administrative Science Quarterly, Journal of Applied Psychology, Journal of Management, Organizational Behavior and Human Decision Processes, Personnel Psychology*, and *Strategic Management Journal.* These journals fairly equally represent the research in micro (e.g., *JAP, PP*), macro (e.g., *ASQ, SMJ*), and mid-range (e.g., *AMJ, JOM, OBHDP*) areas of management. In addition, the leading status of these journals has been well documented in past research (e.g., Coe & Weinstock, 1984; Extejt & Smith, 1990; Johnson & Podsakoff, 1994).

For each journal, we randomly selected 10 articles per volume from the most recent, complete 3 years (1992–1994). This resulted in a total of 210 articles selected (3 years $\times$ 10 articles $\times$ 7 journals). These articles represent approximately 18% of all articles published in the seven journals over these 3 years, and include 26,471 statistical tests. Like past power research (e.g., Mazen et al., 1987a), this number of tests includes all $r$s in correlation matrices. The articles were screened to include only those containing statistical tests for which retrospective power analyses could be conducted. By a considerable margin, our selection of the 3 most recent years' articles covers more years, more journals, and more statistical tests than any previously published assessment of power in management studies.

*Power of statistical tests.* The power was determined for each of the statistical tests reported in the articles. There are three co-determinants of statistical power (Cohen, 1988): The significance criterion ($\alpha$) which is the researcher's long-term probability of erroneously rejecting the null hypotheses, the sample size, and the effect size which reflects the magnitude of a phenomenon in a population (e.g., the impact of the independent variable on the dependent variable or the standardized mean difference). As Mazen et al. (1987a) contended, "The three parameters and statistical power are so closely related that when any three are known, the fourth can be precisely determined" (p. 370). In the present study, Cohen's (1988) power analysis tables and formulae were used to compute statistical power for all forms of correlation, multiple regression, ANOVA, $t$ test, and chi-square (including LISREL) tests. We assumed

TABLE 1

*Frequency and Ascending Cumulative Percentage Distribution of Statistical Power**

| | Small effect size | | Medium effect size | | Large effect size | |
|---|---|---|---|---|---|---|
| Power | Frequencies | Cumulative percent | Frequencies | Cumulative percent | Frequencies | Cumulative percent |
| .99–1.00 | 9 | 4 | 49 | 23 | 124 | 59 |
| .95–.98 | 3 | 7 | 15 | 30 | 22 | 70 |
| .90–.94 | 1 | 7 | 21 | 40 | 17 | 78 |
| .80–.89 | 5 | 9 | 24 | 52 | 21 | 88 |
| .70–.79 | 1 | 10 | 20 | 61 | 8 | 91 |
| .60–.69 | 6 | 13 | 19 | 70 | 10 | 96 |
| .50–.59 | 7 | 16 | 21 | 80 | 4 | 98 |
| .40–.49 | 9 | 20 | 18 | 89 | 1 | 98 |
| .30–.39 | 15 | 27 | 8 | 93 | 2 | 99 |
| .20–.29 | 43 | 47 | 11 | 98 | 0 | 99 |
| .10–.19 | 75 | 83 | 3 | 99 | 1 | 100 |
| .05–.09 | 36 | 100 | 1 | 100 | 0 | 100 |

*Note*: Power assessments assumed .05 alpha and two-tailed tests.
* Frequencies above the dotted line achieved the conventional power of .80.

the use of .05 alpha levels and two-tailed tests. Following past research, we examined the power of global model tests, not individual parameter estimates. As such, our findings represent upper bound estimates of power levels, because individual parameter tests are typically less powerful (Cascio, Valenzi, & Silbey, 1980; Cascio & Zedeck, 1983; Cohen, 1977).

In past power research (e.g., Chase & Chase, 1976; Cohen, 1962; Mazen et al., 1987a), for each article the average statistical power was calculated for small, medium, and large effect sizes. With this conventional procedure, no matter how many tests were contained in a particular article, all articles counted equally in the analysis. In our study, we likewise distributed the mean values for power at each of the three levels of effect size and determined their central tendency measures. Because effect sizes vary by type of statistic, Cohen's (1988) conventional definitions of small, medium, and large effect sizes for each statistic were used. As an example, small, medium, and large effects for Pearson correlations ($r$) are .10, .30, and .50, respectively. Also following Mazen et al. (1987a), we omitted secondary tests like manipulation checks and peripheral reliability estimates.

*Results*. The primary analysis of statistical power in these 210 articles revealed that for small effect sizes, average power across tests was .27; for medium effect sizes, average power across tests was .74; and for large effect sizes, power was .92. More important than the average power, however, were the frequencies of articles reaching or exceeding the recommended power of .80 (see Table 1). For small effect sizes, fewer than

TABLE 2

*Means and Standard Deviations of Statistical Power for Small, Medium, and Large Effect Sizes by Type of Statistic*

| | | Effect size | | |
|---|---|---|---|---|
| Type of statistic | $N$ of analyses | Small | Medium | Large |
| ANOVA | 126 | .22<br>(.21) | .62<br>(.31) | .93<br>(.24) |
| Chi-Square | 113 | .25<br>(.27) | .72<br>(.28) | .93<br>(.16) |
| Correlation | 225 | .33<br>(.29) | .83<br>(.22) | .97<br>(.08) |
| Multiple regression | 165 | .23<br>(.26) | .73<br>(.31) | .92<br>(.18) |
| $T$ test | 17 | .33<br>(.26) | .80<br>(.21) | .95<br>(.17) |

*Note*: Standard deviations are in parentheses. Power assessments assumed .05 alpha and two-tailed tests.

10% of the articles reached or exceeded .80; for medium effect sizes, approximately 52% of the articles reached or exceeded .80; and for large effect sizes, approximately 88% of the articles reached or exceeded .80. Across all effect sizes, the recommended level of power was achieved in only about half the articles, and for small and medium effect sizes combined, the recommended power was achieved in only about 1 in 3 studies (i.e., approximately 37% of the time). This suggests that, across all effect sizes, researchers assume an average risk of sustaining false null hypotheses about half the time, and this risk greatly increases when small and medium effect sizes are considered.

In the next section, power levels are examined with respect to type of statistic, journal, and year of publication. From Cohen's (1988) text, the average statistical power of these tests was approximately .95 (range = .76–1.00). First, analyses of variance (ANOVAs) of power level by type of statistic used in each of the 646 analyses (e.g., ANOVA, chi-square, correlation, multiple regression, and $t$ test) revealed significant main effect differences for small ($d.f.$ = 4, 641; $F$ = 5.70, $p \leq .001$), medium ($d.f.$ = 4, 641; $F$ = 11.35, $p \leq .001$), and large ($d.f.$ = 4, 641; $F$ = 15.96, $p \leq .001$) effect sizes. Table 2 contains descriptive information concerning power levels for each statistic and effect size. Tukey's studentized range tests (an ANOVA procedure) were conducted to determine which statistics had significantly different power levels, at the $p \leq .05$ level. The purpose of this analysis is to determine whether there are significant differences in the actual power level of the various types of statistics commonly employed in the leading journals.

TABLE 3

*Means and Standard Deviations of Statistical Power for Small, Medium, and Large Effect Sizes by Journal*

| Journal | N of analyses | Effect size Small | Medium | Large |
|---|---|---|---|---|
| Academy of Management Journal | 70 | .20<br>(.19) | .62<br>(.32) | .93<br>(.21) |
| Administrative Science Quarterly | 104 | .32<br>(.32) | .80<br>(.27) | .95<br>(.13) |
| Journal of Applied Psychology | 100 | .35<br>(.31) | .82<br>(.25) | .95<br>(.15) |
| Journal of Management | 77 | .33<br>(.31) | .82<br>(.26) | .96<br>(.12) |
| Organizational Behavior and Human Decision Processes | 113 | .17<br>(.18) | .60<br>(.27) | .87<br>(.21) |
| Personnel Psychology | 105 | .30<br>(.25) | .83<br>(.23) | .97<br>(.09) |
| Strategic Management Journal | 77 | .18<br>(.20) | .63<br>(.31) | .87<br>(.22) |

*Note*: *Standard deviations are in parentheses. Power assessments assumed .05 alpha and two-tailed tests.

For both small and medium effects, the power of correlations is not significantly different than that of *t* tests, but is significantly greater than the power of chi-square, multiple regression, and ANOVA. In addition, for medium effects, the power of chi-square and multiple regression are both significantly greater than the power of ANOVA. For large effects, the power of correlation is not significantly different than the power of *t* test or chi-square, but is significantly greater than the power of multiple regression and ANOVA. Further, the power of *t* test, chi-square, and multiple regression are not significantly different, but are all significantly greater than the power of ANOVA.

Second, ANOVAs of power level by journal (e.g., *AMJ*, *ASQ*, *JAP*, etc.) also revealed significant main effect differences for small ($d.f. = 6, 639$; $F = 8.41, p \leq .001$), medium ($d.f. = 6, 639$; $F = 13.00, p \leq .001$), and large ($d.f. = 6, 639$; $F = 7.31, p \leq .001$) effect sizes. Table 3 contains descriptive information concerning power levels for each journal and effect size. Tukey's studentized range tests were conducted to determine which journals had significantly different power levels, at the $p \leq .05$ level. For all effect sizes, the power levels found in *JAP*, *JOM*, *ASQ*, and *PP* were not significantly different. For both small and medium effects, these four journals had significantly greater power than that found in *AMJ*, *SMJ*, and *OBHDP*. For large effect sizes, the power level of *PP* was significantly greater than those in *AMJ*; and *PP*, *JOM*, *JAP*, and *ASQ* had significantly higher power levels than *SMJ* and *OBHDP*.

Third, ANOVAs of small ($d.f. = 2, 637$; $F = .45$, *ns*), medium ($d.f. = 2, 637$; $F = .76$, *ns*), and large ($d.f. = 2, 637$; $F = 1.21$, *ns*) effect sizes by year revealed no significant differences in power levels across the 3 years.

*Studies 2 and 3: Practices and Perceptions of Researchers*

In the prior section, the actual power of small, medium, and large effect sizes, given sample sizes and types of statistics employed, was evaluated across seven leading journals. Yet, that examination does not consider the possibility that researchers actually conduct power analyses, but that other factors might account for the relatively low power levels. For example, it may be that power analyses occur more frequently today, but pragmatic limitations (e.g., budgetary or time constraints) prohibit the increased power called for by past researchers (e.g., Cohen, 1992; Mazen at al., 1987b). Hence, we next examined the perceptions and practices of researchers publishing in leading management journals. Specifically, we sought to understand whether power analyses are undertaken and the reasons for using or not using power analyses.

*Sample and procedure.* We drew from two different but overlapping populations of researchers. First, for Study 2 we surveyed the lead authors of the 210 empirical articles described in our assessment of statistical power in current leading management research (Study 1). Together, Studies 1 and 2 provide information on the actual power level of 210 recent articles, matched with the perceptions and power analysis usage of many of these articles' authors. Then, for Study 3 we examined perceptions and power analysis usage of researchers using multiple regression/correlation analyses, the most commonly used analytic technique in management (e.g., Podsakoff & Dalton, 1987; Weinzimmer et al., 1994).

Of the 210 lead authors identified for Study 2, 29 had multiple publications or overlapped with lead authors for Study 3, reducing the sample to 181. For Study 3, we identified the lead authors of all empirical articles containing regression-based analyses from five journals (*AMJ, ASQ, JAP, JOM,* and *SMJ*) across 4 years (1989–1992). We use Cohen and Cohen's (1983) definition of regression-based models, which includes simple, multiple and moderated linear regression, and all forms of ANOVA. Of the 727 articles published in these journals from 1989 to 1992, 618 (85.0%) were empirical. Of these, 411 (66.5%) used regression-based analyses. Multiple publications across the journals by 42 authors and incomplete or unavailable addresses for another 25 authors reduced the number of authors to 344 (304 domestic; 40 international). Because identical surveys were used and response rates, respondent characteristics, and results were almost identical across both populations, the reporting of remaining procedures and results are combined for Studies 2 and 3.

All 525 lead authors were sent a survey on perceptions and usage of statistical power analysis. An accompanying cover letter explained the study's purpose and participants' anonymity. Approximately 3 weeks later a follow-up reminder letter was mailed. Usable responses were received from 169 authors; another 33 surveys were returned due to authors no longer residing at their listed address. This resulted in a 34.4% response rate which, considering the personal, somewhat private nature of the research, seems fairly reasonable.

The survey was divided into four sections. The first section examined authors' actual usage of statistical power. In the second section, authors using power analysis were asked the extent to which they agreed or disagreed (5-point Likert-type scales, 5 = *strongly agree*) with different reasons for using power analysis. Also for authors using power analyses, a third section was concerned with perceptions of sufficient power levels, sources of effect size, and software programs or sources for power calculations. A final section of the survey, answered by authors not using power analysis, examined reasons for not using power analysis. Demographic information was also collected.

*Respondent characteristics.* Of the 169 respondents, 51 (30.2%) designated organizational behavior as their primary research area; 35 (20.7%) indicated I/O psychology; 23 (13.6%) indicated strategy/policy; 19 (11.2%) indicated human resource management; 19 (11.2%) indicated organizational theory; and 22 (13.0%) indicated other primary research area (e.g., accounting, healthcare, etc.). As anticipated, almost 95% of the respondents held academic positions, with the remainder working for consulting firms, government, or other employers. The average number of research publications over the past 3 years was 7.66 (median = 5.5; $SD$ = 7.4), indicating that this sample can be considered well published. Because the authors were selected from among those publishing in leading management journals, their research behaviors may be considered among the best practiced, compared to those publishing in less acclaimed journals or those not publishing at all.

*Results.* Concerning actual power analysis usage in research projects over the past 3 years, 108 (63.9% of respondents) authors reported never using power analysis. In the rest of this section, the perceptions and practices of the remaining 61 authors who used power analysis in at least one project are discussed. Of these researchers, 94.8% reported a desired power level of .80 or greater in their power calculations. 54.2%

report actually estimating effect sizes all the time, whereas 39.0% report actually estimating effect sizes half the time or less. Of the power analysis users, 51.3% reported using different estimates of effect size when different analytic techniques were used (which is appropriate). Finally, 50.0% of those using power reported never performing post-hoc power analyses when non-significant results were obtained, 17.2% reported performing post-hoc power analyses all the time, and 32.8% reported performing post-hoc power analyses between 10% and 90% of the time.

*Reasons for using power analysis.* Of those using power analyses, 83.6% strongly or moderately agreed (correctly) with the statement, "Sufficient statistical power increases the likelihood that if significant effect sizes exist, they will be detected as statistically significant," and 13.1% disagreed moderately or strongly with this statement. Considerably fewer, 58.3%, moderately or strongly agreed (correctly) with the statement, "Sufficient statistical power ensures that potentially productive research findings are not prematurely terminated," whereas 18.3% strongly or moderately disagreed with this statement. Finally, only 16.2% moderately or strongly agreed with the reason that "Sufficient statistical power ensures that statistically significant results will be obtained," whereas 72.5% moderately or strongly disagreed with this statement. It can be noted that power alone does not *ensure* significance; sizable effects must first exist. Wilcoxon matched-pairs tests of differences between ranks for the first and second reasons ($Z = 1.68, p \leq .05$) and the second and third reasons ($Z = 3.93, p \leq .01$) reveal significant differences.

Concerning the perception of generally sufficient power, almost 90% of those using power analyses consider .80 or greater appropriate. The majority (56.3%) derive effect size estimates from prior research (their own or others), 29.2% get effect sizes from Cohen (e.g., 1988), and 14.6% use other sources. Of those conducting power analyses, a large majority (74.1%) use Cohen's (1977 or 1988) text as their source for power calculations, 11.1% use SPSS software (e.g., SPSS, Inc., 1988), and the remaining 14.9% use other sources.

*Reasons for not using power analysis.* Almost 64% of the respondents reported never using power analyses. Most prominently, 68.9% of these respondents agreed moderately or strongly that they do not use power because, "Editors or reviewers do not require power analysis." Only 13.0% disagreed moderately or strongly with this statement. Again, Wilcoxon matched-pairs tests were employed to assess differences in successive rankings of reasons given for not using power analysis. Significantly fewer researchers ($Z = 2.86, p \leq .01$), 54.4%, agreed moderately or strongly with the statement "I use large sample sizes in my research."

Only 23.2% disagreed strongly or moderately with this statement. Of comparable importance ($Z = 0.74$, *ns*), with 54.3% agreeing moderately or strongly, was that, "My sample size is fixed, or can't easily be changed." There were 28.5% who disagreed strongly or moderately with this statement. The fourth most common reason ($Z = 1.02$, *ns*), with 40.7% agreeing strongly or moderately, was that, "I have limited access to power analysis software programs" (36.3% strongly or moderately disagreed). A significantly smaller portion ($Z = 2.07$, $p \leq .01$), 28.7% of those not using power analyses, agreed strongly or moderately with the statement, "I am not familiar with tools to perform power analysis," although 48.2% disagreed moderately or strongly with this statement. And finally, significantly fewer ($Z = 2.52$, $p \leq .01$), 14.9%, agreed strongly or moderately with the statement, "I am not familiar with power analysis." A large majority (67.4%) disagreed strongly or moderately with this statement, thereby indicating a high awareness of the concept, at least, of power analysis. In response to an open-ended question, the majority (74.1%) indicated there were no other reasons for their not using power analysis. Of the other reasons that were cited, the most frequently mentioned was due to the use of case or qualitative studies ($n = 5$) or because the entire population was sampled ($n = 4$).

## *Discussion*

This paper reports evidence obtained from three examinations of statistical power analysis levels, perceptions, and practices in applied psychology and management research. The first examination (i.e., Study 1) of the actual statistical power levels across a current sample of leading journals corroborates earlier research, suggesting that the average statistical power level in our field is still fairly low. Because we examined only global model tests, not individual parameters, our findings likely overstate the actual power of the tests. As Cohen (1977, pp. 117–119) discusses, due to shared variance across parameter estimates, effect sizes of individual parameter estimates typically will be smaller, resulting in lower power and larger required sample sizes. These findings suggests that, as researchers, we should be more aware of and concerned with statistical power.

Study 1 also examined differences in power levels by year of publication, type of statistic, and journal. There were no significant differences concerning year of publication. Regarding type of statistic, ANOVA was found to have significantly smaller power levels than other types of statistics. This suggests that researchers who rely extensively on ANOVA may want to use larger sample sizes or they will face a greater risk of overlooking significant effects. Of the seven leading journals examined, *Personnel*

*Psychology* contained the highest power levels for medium and large effects, and was among the most powerful for small effects. We believe that differences in power levels across journals are at least partially attributable to the attention that authors, reviewers and editors give to specific conceptual, methodological, and statistical issues. To verify this contention, we carefully examined instructions for authors across these seven leading journals. Only *PP*, in its "Editorial Checklist for Authors" (Campion, 1993), specifically calls for having and reporting adequate statistical power and for interpreting nonsignificant results accordingly. We would expect that such policies direct more attention to statistical power in all stages of research.

From our Study 1 findings of generally low power across the journals, it might be concluded that the impact of past surveys has been minimal or that calls for greater statistical power levels and usage have gone unheeded. However, we think that before any useful diagnoses concerning power usage can be offered, it is necessary to examine the extent to which researchers use power analyses and their reasons for conducting or not conducting power analyses. Thus, we examined the practices and perceptions of researchers publishing in the leading journals.

Across Studies 2 and 3, of the approximately one-third of the respondents using power analysis in at least one of their studies, most were in agreement with commonly accepted practices and perceptions concerning power usage. For example, most seek power of .80 or greater, vary their use of effect sizes with different statistics, and agree with the statement that the reason for using power is to detect significant effect sizes if they actually exist. It appears that some could vary their estimates of effect size when employing different statistics and more could use post-hoc power analyses to a greater extent.

In both Studies 2 and 3, a larger portion, almost two-thirds, of the respondents never use power analysis, and herein lies cause for greater concern. Most fundamentally, these researchers believe there is little call for greater usage of power analysis by journal editors or reviewers. It is unlikely that editors or reviewers have ever intended to convey the message or impression that power analysis usage is unimportant. However, the silence concerning the use of power analysis by these and other gatekeepers seems to have had a pervasive impact on researchers' beliefs and behaviors. Our belief that researchers are aware of power analysis, but choose not to use it, is corroborated by the high degree of disagreement (approximately two-thirds of the non-power users) with the statement, "I am not familiar with power analysis."

It is our contention that researchers' fundamental lack of perceived need for power analysis, found overwhelmingly in Studies 2 and 3, augments the remaining reasons for not using power analyses. On the one hand, there are pragmatic reasons for not using power analyses in a priori or planning stages of research. For example, perceptions concerning large or fixed sample sizes or beliefs that sample sizes can't easily be changed may reflect the difficulties and realities of acquiring data in organizational research. In addition, having limited access to power packages or being unfamiliar with tools used to perform power analysis are reasonable barriers to pursuing power analysis. Yet, on the other hand, we think that there is understandably little motivation to learn how large a sample is large enough, master how to economically increase sample or effect sizes, gain access to power packages, become familiar with power tools, and so forth, without there being some greater perceived need for increased power usage.

So, this raises at least two central questions: (a) Of what purpose or importance, really, is statistical power? (b) Why should researchers be interested in power usage and seek to increase the power of their research? We think that statistical power analysis and levels are important because of both advancement of knowledge and economic reasons. First, with respect to knowledge advancement, inadequate statistical power by definition translates into an increased likelihood that, even if significant effect sizes exist in the population, they will be overlooked in samples used by management researchers. This suggests that the fairly low power in published research may be an indicator of a much larger body of research that is unpublishable, due to non-attainment of statistically significant results. To the extent that this body of research exists, there are grave implications for advancement of knowledge, because this suggests that research actually conducted may not accurately reflect the nature of the phenomenon under investigation. Likewise, and of greater concern, to the extent that underpowered research is published and contains incorrectly accepted false null hypotheses, research can be misleading or contradictory at best and, at worst, highly erroneous (cf. Tversky & Kahnman, 1971). For example, our findings indicate that across all effect sizes, researchers assume an average risk of sustaining false null hypotheses about half the time, and that small and medium effect sizes increase this risk considerably. Because Cohen (1988) maintains that small effect sizes are the norm in social and behavioral sciences, the risks of making Type II errors appear to be quite prevalent.

Another consequence of disregarding statistical power, that of having overpowered samples and tests, can also affect the development and advancement of knowledge. Excessively large samples likely occur less frequently than underpowered samples but still raise a serious concern, particularly in instances of relatively unlimited data access and collection (e.g., PIMS-type databases, secondary data sets on CD-ROM). The risk encountered here is one of oversensitivity to trivial or irrelevant findings (Chase & Tucker, 1976; Lykken, 1968). In any practical sense such findings hold little meaning (i.e., explain little variance), but by virtue of large sample sizes may attain the hallowed status of "statistically significant." Given our field's predilection toward publishing statistically significant findings (Sauley & Bedeian, 1989), as compared to the magnitude or effect size of the phenomenon under investigation, the goals of scientific inquiry are compromised and knowledge is not necessarily advanced.

A second argument for paying greater attention to statistical power concerns economic or efficiency reasons. With insufficient statistical power, there exists an increased likelihood that researchers' time, effort, and other resources are unnecessarily wasted on research endeavors. Simply put, sufficient statistical power allows researchers to more confidently conclude that there are no significant differences (i.e., to more accurately detect and reject false null hypotheses). Without such confidence, researchers may incorrectly conclude that their hypotheses are unsupported by the data or that their data collection efforts have been in vain. Additionally, during planning stages, power analysis allows researchers to more confidently assert that their sample sizes are adequate, thereby saving time and effort on unnecessary (i.e., overpowered) data collection. Finally, post-hoc or evaluative power analyses can be useful for learning whether additional data collection is necessary and, if so, the nature of data to be collected. Post-hoc power tests may also indicate the need for increased effect sizes or alternative measurements and research designs.

To remedy the low power that exists in management journals, we recommend several actions. First, journal editors, reviewers, and other gatekeepers could make more explicit the need to conduct and report power analyses. In instances where statistically nonsignificant results appear, it should be stressed that post-hoc power analyses should reach or exceed the conventional .80 level. Without such power levels, researchers have not given their null hypotheses a fair chance of being rejected and, consequently, spuriously negative results are more likely to occur. From the few (14) articles in our sample of 210 reporting the use of power analyses, 5 used post-hoc power analyses to adjust alpha levels or to explain their nonsignificant findings (e.g., Collins, Hatcher, & Ross, 1993; Cortina, Doherty, Schmitt, Kaufman, & Smith, 1992). More research would seemingly benefit from such practice.

Second, more emphases need to be placed by both editorial gatekeepers and researchers on the role of effect sizes, and less on the single criterion of statistical significance ($\alpha$, or $p$), for judging the findings of results. A number of researchers, including Cohen (1988, 1992), Hunter and Schmidt (1990), Lykken (1968), and Rosenthal (1991) have articulated the advantages of reporting and evaluating effect sizes in research. In short, the reporting of effect sizes allows results to be evaluated in terms of their practical significance and increases the likelihood that effects (i.e., magnitude of a phenomenon in a population or impact of independent variables) can be compared across studies. The reporting of effect sizes therefore increases the replicability of research streams and the ability to aggregate effect sizes through meta-analyses. However, neither replication nor meta-analyses negate the need for individual researchers to fully utilize statistical power analyses. Indeed, statistically underpowered research may never overcome what Rosenthal (1979) described as the "file drawer problem," meaning that due to a field's obsession with statistical significance, a study's nonsignificant findings may relegate it to terminal existence in a file drawer.

Finally, researchers should realize that they have everything to gain, and little to lose, by examining the statistical power in their research more carefully. Perhaps by more clearly recognizing the consequences of low power, researchers will see the personal (e.g., economic, efficiency) gains and the potential impact on the advancement of knowledge. Likewise, greater sensitivity to the squandered time, effort, and financial resources devoted to unnecessarily large data collections should provide further incentive to investigate appropriate power before conducting research. Over a dozen microcomputer software packages now exist for computing power in both planning and retrospective analysis formats (for a review, see Goldstein, 1989). Cohen's (1988) manual on power analysis, complete with straightforward tables for almost all statistical tests, is both comprehensive and easy to use. Other hands-on sources for power analysis, generally based on Cohen's work, include Kraemer and Thiemann (1987), Lipsey (1990), and Rosenthal and Rosnow (1991).

Beyond simply addressing sample size considerations, these and other sources provide numerous suggestions for increasing effect size and statistical power. One suggestion involves improving methodological techniques and employing better measurement models to increase effect sizes (e.g., more reliable measures and stronger tests or manipulations). Measurement unreliability attenuates validity estimates, reduces statistical power, and increases the need for larger samples (Sawyer & Ball, 1981; Schmidt, Hunter, & Urry, 1976; Sutcliffe, 1980). Range restriction is another factor that inhibits effect size estimates, resulting also in lower power (Schmidt et al., 1976). Researchers should correct for measurement unreliability and range restriction, using formulae suggested by Schmidt et al. (1976) or Schmitt and Klimoski (1991), and they should also attempt to improve actual measurement reliability and avoid range restriction in sampling. As another suggestion, more sensitive research designs can be employed (e.g., blocking, covariates, within-subject designs) to reduce unexplained error (Greenwald, 1976; Rosenthal & Rosnow, 1991). Finally, more powerful statistical techniques (e.g., factorial, repeated measures, combined predictors, non-parametric statistics) and adjustments to alpha can be used (Cascio & Zedeck, 1983; Collins et al., 1993; Sawyer & Ball, 1981). Cascio and Zedeck (1983), for example, provide detailed treatment of how to adjust alpha (Type I error risk) to reduce Type II errors, identifying circumstances (e.g., fixed sample, anticipated effect size identified) when it is appropriate to adjust alpha to maximize statistical power.

In sum, our survey of current leading journals indicates that statistical power levels in applied psychology and management are low. Findings from our two surveys of researchers indicate that less than a third of the researchers typically conduct statistical power analyses, whereas approximately two-thirds report never using power analyses. The reasons for not conducting power are not ones of non-awareness, but are more related to the belief that statistical power analyses are not required by journal editors or reviewers. We hope that our investigation increases researchers' and gatekeepers' sensitivity to the role of and need for statistical power, thereby resulting in increased power levels.

## REFERENCES

Campion MA. (1993). Article review checklist: A criterion checklist for reviewing research articles in applied psychology. PERSONNEL PSYCHOLOGY, *46*, 705–718.

Cascio WF, Valenzi ER, Silbey V. (1980). More on validation and statistical power. *Journal of Applied Psychology, 65*, 135–138.

Cascio WF, Zedeck S. (1983). Open a new window in rational research planning: Adjust alpha to maximize statistical power. PERSONNEL PSYCHOLOGY, *36*, 517–526.

Chase LJ, Chase RB. (1976). A statistical power analysis of applied psychological research. *Journal of Applied Psychology, 61*, 234–237.

Chase LJ, Tucker RK. (1975). A power-analytic examination of contemporary communication research. *Speech Monographs, 42*, 29–41.

Chase LJ, Tucker RK. (1976). Statistical power: Derivation, development, and data-analytic implications. *The Psychological Record, 26*, 473–486.

Coe R, Weinstock I. (1984). Evaluating the management journals: A second look. *Academy of Management Journal, 27*, 660–665.

Cohen J. (1962). The statistical power of abnormal-social psychological research: A review. *Journal of Abnormal and Social Psychology, 65*, 145–153.

Cohen J. (1977). *Statistical power for the behavioral sciences* (rev. ed.). New York: Academic Press.

Cohen J. (1988). *Statistical power for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Cohen J. (1992). Statistical power analysis. *Current Directions in Psychological Science, 1*, 98–101.

Cohen J, Cohen P. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Collins D, Hatcher L, Ross TL. (1993). The decision to implement gainsharing: The role of work climate, expected outcomes, and union status. PERSONNEL PSYCHOLOGY, *46*, 77–104.

Cortina JM, Doherty ML, Schmitt N, Kaufman G, Smith RG. (1992). The "big five" personality factors in the IPI and the MMPI: Predictors of police performance. PERSONNEL PSYCHOLOGY, *45*, 119–140.

Extejt MM, Smith JE. (1990). The behavioral sciences and management: An evaluation of relevant journals. *Journal of Management, 16*, 539–551.

Goldstein R. (1989). Power and sample size via MS/PC-DOS computers. *The American Statistician, 43*, 253–260.

Greenwald AG. (1976). Within-subject designs: To use or not to use? *Psychological Bulletin, 83*, 314–320.

Hunter JE, Schmidt FL. (1990). *Methods of meta-analysis: Correcting error and bias in research findings.* Newbury Park, CA: Sage.

Jick TD. (1979). Mixing qualitative and quantitative methods: Triangulation in action. *Administrative Science Quarterly, 24*, 602–611.

Johnson JL, Podsakoff PM. (1994). Journal influence in the field of management: An analysis using Salancik's index in a dependency network. *Academy of Management Journal, 37*, 1392–1407.

Kraemer HC, Thiemann S. (1987). *How many subjects? Statistical power analysis in research.* Newbury Park, CA: Sage.

Lipsey MW. (1990). *Design sensitivity: Statistical power for experimental research.* Newbury Park, CA: Sage.

Lykken DT. (1968). Statistical significance in psychological research. *Psychological Bulletin, 70*, 151–159.

Mazen AM, Graf LA, Kellogg CE, Hemmasi M. (1987a). Statistical power in contemporary management research. *Academy of Management Journal, 30*, 369–380.

Mazen AM, Hemmasi M, Lewis MF. (1987b). Assessment of statistical power in contemporary strategy research. *Strategic Management Journal, 8*, 403–410.

Mone MA, McKinley W. (1993). The uniqueness value and its consequences for organization studies. *Journal of Management Inquiry, 2*, 284–296.

Neuliep JW. (1991). *Replication research in the social sciences.* Newbury Park, CA: Sage.

Orr JM, Sackett PR, Dubois CLZ. (1991). Outlier detection and treatment in I/O psychology: A survey of researcher beliefs and an empirical illustration. PERSONNEL PSYCHOLOGY, *44*, 473–486.

Podsakoff PM, Dalton DR. (1987). Research methodology in organizational studies. *Journal of Management, 13*, 419–441.

Rosenthal R. (1979). The "file drawer problem" and tolerance for null results. *Psychological Bulletin, 86*, 638–641.

Rosenthal R. (1990). Replication in behavioral research. *Journal of Social Behavior and Personality, 5*, 1–30.

Rosenthal R. (1991). *Meta-analytic procedures for social research* (rev. ed.). Beverly Hills, CA: Sage.

Rosenthal R, Rosnow RL. (1991). *Essentials of behavioral research: Methods and data analysis* (2nd ed.). New York: McGraw-Hill.

Rossi JS. (1990). Statistical power of psychological research: What have we gained in 20 years? *Journal of Consulting and Clinical Psychology, 58*, 646–656.

Sauley KS, Bedeian AG. (1989). .05: A case of the tail wagging the distribution. *Journal of Management, 15*, 335–344.

Sawyer A, Ball A. (1981). Statistical power and effect size in marketing research. *Journal of Marketing Research, 18,* 275–290.

Schmidt FL, Hunter JE, Urry VW. (1976). Statistical power in criterion-related validity studies. *Journal of Applied Psychology, 61,* 473–485.

Schmitt NW, Klimoski RJ. (1991). *Research methods in human resource management.* Cincinnati: South-Western.

Sedlmeier P, Gigerenzer G. (1989). Do studies of statistical power have an effect on the power of studies? *Psychological Bulletin, 105,* 309–316.

SPSS, Inc. (1988). *SPSS-X user's guide* (3rd ed.). Chicago: Author.

Stone-Romero EF, Anderson LE. (1994). Relative power of moderated multiple regression and the comparison of subgroup correlation coefficients for detecting moderating effects. *Journal of Applied Psychology, 79,* 354–359.

Sutcliffe JP. (1980). On the relationship of reliability to statistical power. *Psychological Bulletin, 88,* 509–515.

Tversky A, Kahnman D. (1971). Belief in the law of small numbers. *Psychological Bulletin, 76,* 105–110.

Weinzimmer LG, Mone MA, Alwan LC. (1994). An examination of perceptions and usage of regression diagnostics in organization studies. *Journal of Management, 20,* 179–192.