Strategic Management Journal, Vol. 8, 403–410 (1987)

RESEARCH NOTES AND COMMUNICATIONS

ASSESSMENT OF STATISTICAL POWER IN CONTEMPORARY STRATEGY RESEARCH

A. MAGID M. MAZEN, MASOUD HEMMASI AND MARY FRANCES LEWIS,
College of Business, Illinois State University, Normal, Illinois, U.S.A.

The concept of statistical power was reviewed, and the power of 44 recently published empirical studies in strategic management was analyzed. Using small, medium, and large estimates of effect size, standardized 0.05 α, and assuming nondirectional nulls, the mean power figures were 0.23, 0.59, and 0.83 for the three levels, respectively. These results were generally similar to findings in other social sciences and were considered particularly important for strategic management research, given the correlational nature of many strategic management investigations and the complexity and evolving stage of the field which make small effect size more likely. Ways to improve statistical power in strategic management research were discussed.

INTRODUCTION

The average person would probably neither use a hammer to kill a mosquito, nor a candle to search for a needle on a football field at night; the former is too powerful a means, and the latter is just too weak. Although the concept of statistical power is basically similar to the above analogies, it is not as obvious and often causes confusion. In fact, surveys of sample research in abnormal-social psychology (Cohen, 1962), education (Brewer, 1972), communication (Chase and Tucker, 1975), applied psychology (Chase and Chase, 1976), and marketing (Sawyer and Ball, 1981) have all concluded that research in these areas lacks statistical power. That is, the probability of rejecting the null hypotheses in these areas is much lower than acceptable conventional levels.

These findings pose a question to researchers in strategic management: Does our research have adequate statistical power? The present study addresses this question by, first, providing a brief review of the determinants of statistical power; second, assessing the power of a sample of contemporary strategic management research; and, third, highlighting the major ways to improve power in strategic management research using examples from the sample.

STATISTICAL POWER

Table 1 provides definitions of statistical power (1–β) and its three main determinants: significant level (α), effect size (ES), and sample size (n) (cf. Cohen, 1977: 4–14). The four power parameters are so related that when any three of them are fixed, the fourth can be completely determined. Concretely, inasmuch as there is an inverse relationship between Type I error (erroneously concluding the presence of a phenomenon in a population) and Type II error (mistakenly sustaining the null), an increase in α decreases β, and hence increases power for any given sample size. Similarly, the larger the sample the smaller the error, and hence the more accurately the phenomenon under investigation

0143–2095/87/040403–08$05.00

Received 18 February 1985
Revised 13 October 1986

Table 1. Parameters of statistical power

1. Statistical power (1-β): the *a priori* probability of rejecting H_0. A test is powerful if, when wearing the 'glasses' of the test, one can infer from a sample what truly exists in its population.
2. The significance criterion (α): the expression of the researcher's *policy* with regard to risking the mistaken rejection of H_0. Thus α is a long-term error rate for rejecting when H_0 is true.
3. Effect size: the degree to which the phenomenon is present in the population (i.e. the degree to which the 'null' hypothesis is not really null).
4. Sample size (*n*): the number of observations per test. Optimally one should specify α, effect size, and desired power and then solve for *n*. In this form power, like α, becomes a matter of policy; it is to be viewed in regard to *long-term* probability of Type II error (β). Generally, however, *n* is a codeterminant of power.

can be represented, leading to higher probabilities of rejecting a false null hypothesis. Thus, the larger the sample, the smaller the effect it can detect. Alternatively, the smaller the effect of a particular phenomenon in the population, everything else being equal, the larger the sample needed to discover its weak signal relative to irrelevant noise. Finally *ceteris paribus*, as effect size of a phenomenon increases, the power to discover it, and hence reject the null hypothesis, increases.

Ideally, the researcher should first decide on the power level (1–β) in the planning stage of the study by deciding on the acceptable risk of Type II error (β). Based on the general notion that failure to find is less serious than finding what does not exist, a position which accords with the conventional scientific view, Cohen (1977) suggested that Type I error is four times as serious as Type II. Because the risk of the former is conventionally set at α = 0.05, Cohen offered 0.20 as a conventional Type II error rate (β), setting conventional power at 0.80. Obviously, then, the cost of committing *both* Type I and Type II errors should always be considered in deciding on the desirable level of power.

Where power is not planned *a priori*, it should be estimated *a posteriori* via its three codeterminants, α, *n* and effect size. Of these, the magnitude of effect size is perhaps the least obvious. One way to estimate effect size for a specified topic is via Cohen's (1977) formulae which are based on the amount of variance typically explained in well-conceived research on the subject. In cases where specific effect size levels are difficult to determine (e.g. in new research areas), three conventional effect size levels can be used (Cohen, 1977). These represent 'small', 'medium', and 'large' effect size corresponding to possible sizes of the phenomena in the population. For example, when the difference between two means reflects small effect size, the *non*overlapped area of the distributions of the two populations is about 15 percent. Correspondingly, medium and large effect size reflect 30 percent and 50 percent nonoverlap, respectively. Equipped with α, *n*, and effect size, the researcher can then use power tables available (cf. Cohen, 1977; Cohen and Cohen, 1983) for the statistical tests (i.e., R^2, β, *F*, *t*, *r*, χ^2, and *P*) commonly associated with the most frequently utilized designs.

METHOD

Previous power analytical works in psychology, education, and marketing have all investigated empirical research in one volume of one journal in the respective discipline. Because the ratio of conceptual to empirical articles in the *Strategic Management Journal (SMJ)* is relatively high, the statistical power of research in the 1982, 1983, and 1984 volumes of *SMJ* as well as strategy research in the 1984 volume of the *Academy of Management Journal (AMJ)* was analyzed in the present study. In keeping with previous practice, the article served as the unit of analysis and only major significance tests of hypotheses for which power tables are available were examined; secondary tests such as manipulation checks and peripheral reliability estimates were omitted.

For each test of significance, Cohen's three conventional levels of effect size (small, medium, and large) were adopted. When cell sizes were unequal, harmonized mean functions were used. Also, when factorial and complex designs of analysis of variance were employed, or when interaction was considered, *n* for the respective factor or interaction was determined according to Cohen's (1977) formulae. An $\alpha = 0.05$ and the nondirectional version of the null hypothesis were used uniformly. Once power estimates for all tests in a study were computed, the average statistical power for each article was calculated for small, medium, and large effect size. By this conventional procedure, no matter how many tests were involved in a particular study, all articles counted equally in the description of the volumes examined. The mean power value of the studies at each of the three effect size levels were then distributed and their central tendency measures determined. The procedure described above is consistent with previous surveys of statistical power analysis.

RESULTS

Of the 81 articles in the 1982–84 volumes of *SMJ*, 28 were empirical, contained major significance tests for which power tables exist, and had detectable sample size information for those tests. In addition, 16 articles in the 1984 volume of *AMJ* also qualified, bringing the total number of articles examined in the present study to 44. The total number of tests in the articles was 3665 (1921 in *SMJ* and 1744 in *AMJ*). There were 331 R^2, 1122 β, 268 F, 511 t, 1140 r, 82 χ^2, and 211 tests of proportion. The frequency and cumulative percentage distributions and central tendency measures of power means for the three ES levels are given in Table 2.

The power to detect the size of the three effect levels previously defined, were as follows:

Small effect. On the average, the studies reviewed had about one chance in four of detecting small effects. Of the articles examined, 89 percent had less than a 50–50 chance of detecting significant relationships. Only three articles 7 percent realized the conventional 80 percent power or better. The mean of power estimates for this effect category was noticeably small, 0.23. In fact, this mean was even inflated by several extreme sample points. Thus, if researchers were investigating small effect size, their tests and designs had only 23 percent power

Table 2. Frequency of statistical power in 44 strategy articles for small, medium, and large effects

Small effect			Medium effect			Large effect		
Power	Frequency	Ascending cumulative (%)	Power	Frequency	Ascending cumulative (%)	Power	Frequency	Ascending cumulative (%)
0.99			0.99	6	14	0.99	12	27
0.95–0.98			0.95–0.98	2	18	0.95–0.98	6	41
0.90–0.94	2	5	0.90–0.94	2	23	0.90–0.94	5	52
0.80–0.89	1	7	0.80–0.89	3	30	**0.80–0.89**	6	66
0.70–0.79	1	9	0.70–0.79	4	39	0.70–0.79	7	82
0.60–0.69	1	11	0.60–0.69	4	48	0.60–0.69	3	89
0.50–0.59			**0.50–0.59**	6	61	0.50–0.59	2	93
0.40–0.49	2	16	0.40–0.49	4	71	0.40–0.49	1	96
0.30–0.39	2	21	0.30–0.39	7	86	0.30–0.39		
0.20–0.29[a]	5	32	0.20–0.29	3	93	0.20–0.29	1	98
0.10–0.19	16	68	0.10–0.19	3	100	0.10–0.19	1	100
0.04–0.09	14	100	0.04–0.09			0.04–0.09		
n	44			44			44	
Mean	0.23			0.59			0.83	
Median	0.13			0.53			0.88	
Mode	0.15			0.55			−0.99	

[a] Categories of the three means are shown in bold.
The dotted line is at the 0.80 conventional power level.

to uncover the presence of such effects in the population. Put differently, these studies assumed an average 77 percent risk of erroneously sustaining the null hypothesis.

Medium effect. The picture improves, though not considerably, when one posits medium effects in the population. The mean power for this level was only 0.59, and only 30 percent of the studies achieved the conventional 0.80 power level or better. That is, more than two-thirds of the research surveyed here did not achieve the conventional 0.80 power to capture medium effect sizes.

Large effect. The improvement is quite noticeable here. Sixty-six percent of the articles achieved the 0.80 power level or better, and 27 percent had 0.99 chance of correctly rejecting the null hypothesis. This means that a researcher who investigated, say, large differences between means had an average of more than four chances in five to reject the null hypothesis. But these results also indicate that even when effect size was large, one-third of the articles in the sample were not statistically powerful.

To put these results in perspective, the statistical power of strategic management research was compared to that of different areas of the social sciences (see Table 3). Except for marketing, there is a strong comparability among the power of strategic management research and that of other disciplines. Generally, when effect size was small or medium the average study did not achieve the conventional 0.80 level of statistical power. Only for large effect size was research in these areas powerful enough to detect the phenomena under analysis. A more accurate comparison should of course hold time constant, for it is possible that over the years any or all of the other disciplines have advanced their methodological rigor. However, it is clear that the reports by Brewer (1972) and Chase and Chase (1976) did not differ appreciably from the results reported in the present study or in Cohen's (1962) survey.

To complete the picture we investigated whether statistical power of the present strategic management sample varied (at each of the three effect size levels) along three dimensions: (a) research area, (b) research type, and (c) data source. Table 4 presents power means for the three effect size levels, for selected categories within each of the three dimensions. Simple and multiple regression analyses with two-way interactions, using the three dimensions as sets, showed no significant results.[1] Thus generic guidelines regarding research power could not be developed along these dimensions and categories.

The following section will put the above findings in perspective with respect to the size of effect in, and characteristics of, strategy research.

DISCUSSION AND IMPLICATIONS

From Table 2 it is obvious that the answer to our original question (Is Strategic Management research statistically powerful?) depends on the effect size level typically dealt with in the various areas of the field. If the effect size is large, one can safely conclude that strategic management research is, on the average, statistically powerful. However, the present results show that a similar statement may not be strongly supported regarding medium effect size research and definitely cannot be made if the effect size is small.

An empirical assessment of effect size detected in strategy research can be obtained from meta-analyses of the various areas of the field, and is beyond the scope of this article. However, we seek to suggest that, on the average, effect size in strategy research may be typically small. Our judgement is based on several considerations. First, small effect size is considered the norm, not the exception in the behavioral sciences (cf. Cohen and Cohen, 1983). Even in disciplines such as marketing where statistical power has been found to be relatively high (e.g. Sawyer and Ball, 1981), meta-analyses of research results show effect size to be usually small (e.g. Peterson, Albaum and Beltramini, 1985). The similarity recently drawn between the growth rate of contemporary strategic management and that of marketing in the early 1960s (Schendel, 1985) makes it even more plausible to assume that effect size in strategy research is, similar to that of marketing, small.

Second, because strategies consist of the integration of many dimensions which, in turn, can be configured in endless combinations, any single study must be somewhat circumscribed in

[1] With n = 44, and α = 0.05, these analyses achieved 0.80 power of detecting R^2 = 0.20 (medium effect size) or more.

Table 3. Interdisciplinary comparisons of statistical power for small, medium, and large effect sizes

Research outlet	Sample size No. of articles	No. of tests	Mean of statistical power Small effect	Medium effect	Large effect
Journal Abnormal and Social Psychology (1960)[a]	70	2088	0.18	0.48	0.83
American Educational Research Journal (1969–1970)[b]	47	373	0.14	0.58	0.78
Nine Communication Journal (1974)[c]	46	1298	0.18	0.52	0.79
Journal of Applied Psychology (1974)[d]	121	3373	0.25	0.67	0.86
Journal of Marketing Research (1979)[e]	23		0.41	0.89	0.98
Strategic Management Research (SMJ and *AMJ)*	44	3665	0.23	0.59	0.83

[a]See Cohen (1962).
[b]See Brewer (1972).
[c]See Chase and Tucker (1975).
[d]See Chase and Chase (1976).
[e]See Sawyer and Ball (1981).

its scope (Hambrick, 1984). As such, we may not expect the average study to incorporate all relevant variables, much less to capture large proportions of their variance. Third, strategic management research, by and large, uses correlational techniques; over 90 percent of the articles surveyed here did. The serious design problems (e.g. threats to internal validity) found by Mitchell (1985) in correlational research conducted in organizational settings are not uncommon in

Table 4. Mean power of the 44 articles by selected categories

Category	Effect size Small	Medium	Large	No. of cases
Research area				
Strategy form processes	0.23	0.61	0.83	23
Strategy form elements	0.31	0.76	0.94	3
Strategy implementation processes	0.15	0.50	0.78	5
Strategy implementation elements	0.25	0.51	0.74	5
Strategic management processes	0.18	0.65	0.91	4
Other	0.32	0.65	0.86	4
				44
Research type[a]				
Theory building exploration	0.32	0.68	0.88	13
Theory building concept development				
Theory building H_0: generation	0.31	0.61	0.86	4
Theory testing internal validity	0.15	0.67	0.85	7
Theory testing external validity	0.20	0.53	0.77	23
				47
Data source[a]				
Interview	0.13	0.52	0.82	7
Records and data banks	0.24	0.59	0.80	26
Questionnaire	0.25	0.58	0.82	21
				54

[a]Some articles in this category were classified under more than one subcategory.

strategy research. For example, only 15 percent of the studies in the present sample included information necessary to assess the psychometric properties of their instruments and discussed other issues related to statistical conclusion validity (e.g. cross-validation). Although the absence of such information does not necessarily mean lack of research validity, we took it, conservatively, as an indication of less than adequate attention to research design issues. These design problems can only 'fog' and reduce the amount of variance uncovered in strategic management research. Thus, even when effect size is not small, fallible designs and instruments are likely to detect only a fraction of it. Boruch and Gomez (1977) showed that the impact of unreliable instruments and invalid relationships on the variance explained, and hence on detected effect size is indeed multiplicative. Fourth, many strategic management investigations are plagued with small sample problems, particularly when the firm is the unit of analysis. In the present sample, the median *n* comprised 41 cases, with 30 percent of the studies using less than 27 cases. As was pointed out in the introduction, small *n* has a crippling effect on research power. Taken together, the joint impact of the above factors reduces detectable effect size even further. Finally, if our estimation of effect size is erroneously too conservative, the cost may entail more research insight and rigor. This, in our opinion, is an acceptable price compared to the perpetuation of current practices that may result from entertaining the possibly erroneous assumption that medium or large effects are the norm in strategic management research. In fact, even in the medium effect size category, the present study shows that the risk of committing Type II error in strategic management research was over 40 percent (see Table 2), double the conventionally acceptable β level. We must equally stress, however, that a small effect size does not mean invaluable research. Indeed, not seeking as little as 1 percent effect size may cause important and practical developments in the field to be ignored. Moreover, some large effect size phenomena may be too obvious, or even trivial, to justify scientific interest.

Whatever the size of the effect typically investigated in strategic management research, one should always consider means to address research power via a careful consideration of its codeterminants. For example, one common practice to increase power is to employ a large sample. However, statistical power planning shows that more (cases) may neither be efficient, nor the only route to powerful research. In the remaining part of this note we will focus on two examples from the present sample where methodological issues, including *n*, and theoretical insight may have affected research power; generic and more comprehensive guidelines can be found elsewhere (e.g. Cohen, 1977; Sawyer and Ball, 1981).

Consider an article examining intended length of stay in the firm, among internally promoted and externally hired executives. Like 68 percent of the studies surveyed here, a questionnaire was used in data collection, and similar to 17 percent of the studies, a large *n* (=1178) was generated. However, the scales measuring the independent and dependent variables were one-item each. Despite the large *n*, the researchers had to declare the amount of explained variance 'insufficient for meaningful conclusions'. From a statistical perspective, and given this huge *n*, the power achieved in the study was 0.64, 0.995, or 0.995 for small, medium, and large effect size respectively. But this assumes that the study was methodologically well conceived. With the questionable reliability and validity of one-item scales, we do not really know if there was no difference between the intended length of stay of outsider and insider executives, or if the difference was 'fogged' by unreliable instruments.

Power planning could have improved the situation in the above study appreciably. First, an estimation of effect size in this area of research would have been necessary. The merit of this step is that it links present research to past, and hence provides a systematic accumulation of findings which, in turn, expedites theory development. Assuming effect size for the above study was medium, only 160 responses would have sufficed to achieve 0.80 power, and 300 responses to reach 0.99 power. This huge savings (at least 800 responses) could have been simultaneously directed toward improving the psychometric properties of the instruments and testing additional hypotheses, if any. For example, even if the sampled executives could be measured only once, a small random subgroup could have received a longer questionnaire with multiple measures of the same or multiple constructs.

Or the total sample could have been divided randomly into several groups where each receives a somewhat different combination of items measuring the dependent and independent variables, confounds, validity checks, etc.; anything to improve and assess instruments' fidelity. Further, had the questionnaires been scaled differently, the researcher could have had an even better opportunity to reduce method variance, providing a more realistic estimate of the explained variance and its corresponding effect size (Mitchell, 1985).

Even when small sample is a realistic limitation to strategic management research (as may have been the case with about 30 percent of the articles whose n was less than 27), designs such as repeated measures provide an experimental control with significant impact on power. In the above example, two repeated measures and $\alpha = 0.05$ would have required the responses of only 40 executives (3 percent of the original 1178 sample) to bring power to its conventional 0.80 level. Considering the typical complaint from executives 'bombarded' by research inquiries, the researcher must weigh the relative savings in executives' time against his/her own time invested in power planning (which we roughly estimate at about 1 hour per study to be added to the literature review stage).

Fredrickson's (1984) study provides an example where theoretical insight and methodological rigor may have impacted research power favorably. Working with only 37 firms ($=n$), the researcher developed a model which viewed strategic comprehensiveness as a behaviorally based, decision-making process with four operationalized steps. Compared to the global measures of the construct suggested in traditional models, Fredrickson's model offered an opportunity to maximize detected variance since firms may vary along any or all of the four steps. Additionally, multiple scores were employed to assess the study's constructs, questionnaires included single- and multiple-item scales, potential confounds were assessed and partialled, validity of the theoretical model was checked, and construct validity was assessed. Further, the use of a theoretical model offered the study the benefit a directional (twice as large), α, which increased power. Because the study was well conceived, its large correlations (range from $r = 0.43$ to $r = 0.49$) could be used to approximate the population effect size (Cohen, 1977). With a directional $\alpha = 0.05$ and large effect size, the power of the study exceeded the conventional 0.80 level using a sample size of only 37. Thus while the researcher found numerous significant relationships at the stated α, there was also a low level of risk (β) associated with the study's null conclusions. Unfortunately, even in this well-conceived study no explicit consideration or calculation of power was given.

CONCLUSION

The conclusion of this article is that strategic management researchers need to pay 'formal' attention to statistical power planning, whether they have access to large or small samples. It is worth the time and effort. On the one hand, power assessment enables investigators and consumers of strategic management research to assess the risk of Type II error in every study; on the other, power planning leads to better, cumulative, and more economical research by underscoring the importance of theoretical insight and methodological rigor. Without the incorporation of statistical power planning in research design or its *a posteriori* estimation in published reports, many lines of useful inquiries may be undeservedly abandoned, a costly alternative for the growing field of strategic management.

ACKNOWLEDGEMENTS

The authors wish to thank two anonymous reviewers and Colin Camerer for helpful comments on earlier drafts.

REFERENCES

Boruch, R. and H. Gomez. 'Sensitivity, bias, and theory in impact evaluations', *Professional Psychology*, 1977 (November), pp. 411–434.

Brewer, J. 'On the power of statistical tests in the *American Educational Research Journal*', *American Educational Research Journal*, **9**, 1972, pp. 391–401.

Chase, J. and B. Chase. 'A statistical power analysis of applied psychological research', *Journal of Applied Psychology*, **61**(2), 1976, pp. 234–237.

Chase, L. and R. Tucker. 'A power-analytic examination of contemporary communication research', *Speech Monographs*, **42**, 1975, pp. 29–41.

Cohen, J. 'The statistical power of abnormal-social psychological research: a review', *Journal of Abnormal and Social Psychology*, **65**(3), 1962, pp. 145–153.

Cohen, J. *Statistical Power Analysis for the Behavioral Sciences*, Academic Press, New York, 1977.

Cohen, J. and P. Cohen. *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences* (2nd edn), Lawrence Earlbaum Assciates, Hillsdale, New Jersey, 1983.

Fredrickson, J. 'The comprehensiveness of strategic decision processes: extension, observations, future directions', *Academy of Management Journal*, **27**(3), 1984, pp. 445–466.

Hambrick, D. 'Taxonomic approaches to studying strategy: some conceptual and methodological issues', *Journal of Management*, **10**(1), 1984, pp. 27–41.

Mitchell, T. R. 'An evaluation of the validity of correlational research conducted in organizational settings', *Academy of Management Review*, **10**(2), 1985, pp. 192–205.

Peterson, R., G. Albaum and R. Beltramin. 'A meta-analysis of effect sizes in consumer behavior experiments', *Journal of Consumer Research*, **12** (June), 1985, pp. 97–103.

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

Schendel D. 'Strategic management and strategic marketing: what is strategic about either one?' In Thomas, H. and D. Gardner. (eds), *Strategic Marketing and Management*, Wiley & Sons, New York, 1985, pp. 41–63.