CASE SURVEY METHODOLOGY: QUANTITATIVE ANALYSIS OF PATTERNS ACROSS CASE STUDIES

RIKARD LARSSON
Lund University

Case surveys bridge the gap between nomothetic surveys and idiographic case studies to combine their respective benefits of generalizable, cross-sectional analysis and in-depth, processual analysis. Methodological fragmentation has limited systematic utilization of numerous managerially relevant case studies. This article develops a comprehensive procedure synthesizing the individual strengths of previous efforts and illustrates this method in a case survey of mergers and acquisitions. The empirical application is used to generate directions for future methodological development.

Researchers tend to favor one of two major methods of gathering data, either the nomothetic survey method, which emphasizes quantitative analysis of a few variables across large samples, or the idiographic case study method, which focuses primarily on the qualitative, multiaspect, in-depth study of one or a few cases. Although the two preferences can be traced back to a classical paradigmatic debate (Burrell & Morgan, 1979), the key issue—many issues in few cases versus few issues in many observations—may actually be more pragmatic than paradigmatic. Few would disagree with the desirability of studying many issues in many cases instead of sacrificing either a number of issues or a number of observations or cases (cf. Lazarus, 1971). It is the lack of resources and methods with which to do both that limits the practical choice to one or the other. The present research reviewed and refined a methodology that can transcend these limitations and thereby bridge the nomothetic-idiographic gap (cf. Morey & Luthans, 1984).

A lack of resources with which to conduct a sufficient number of rich case studies for statistical testing can be overcome by using previous case studies (cf. Lee, 1991). Several researchers have attempted to improve the

poor utilization of prior research by developing methods for the quantitative meta-analysis of previous surveys to summarize their findings (e.g., Bangert-Drowns, 1986; Cooper, 1984; Glass, McGaw, & Smith, 1981; Hunter, Schmidt, & Jackson, 1982; Hyman, 1972; Kiecolt & Nathan, 1985). The methodological development of the meta-analysis of case studies has received much less attention. The few existing attempts have been called case survey (Lucas, 1974; Yin & Heald, 1975), case meta-analysis (Bullock & Tubbs, 1987), and structured content analysis of cases (Jauch, Osborn, & Martin, 1980); various nameless versions exist as well (e.g., Miller & Friesen, 1977). These methods differ significantly in strengths and weaknesses but basically represent the same process of quantifying a group of existing case studies for statistical analysis. To avoid confusion, I use "case survey" here since it is the earliest, simplest, and most distinctive of the three names.

The case survey method is one of the most promising approaches for bridging the nomothetic-idiographic research gap and using prior empirical studies. It can overcome the problem of generalizing from a single case study and at the same time provide more in-depth analysis of complex organizational phenomena than questionnaire surveys. The ability of case studies to deal with processual and multiple stakeholder considerations through using longitudinal, multisource data makes them well-suited for management inquiry into unique situations. Their inability to allow statistical examination of patterns across varying situations can be overcome through cross-sectional analysis of sets of cases. Thus, numerous cases studies represent a fertile pool of managerially relevant findings whose individually limited scientific contributions can be enhanced through systematic analysis of patterns across cases.

The fragmentation of previous attempts has hindered the development of a methodological capacity to fulfill this potential. Several weaknesses resulting from lack of review of prior methodological developments and applications offset their strengths. The purpose of this research was to overcome the various weaknesses of earlier attempts by synthesizing their strengths. I reviewed case survey research as a basis for developing an improved, comprehensive statement of the case survey method and applied the method in a study of mergers and acquisitions.

CASE SURVEY METHODOLOGY

The case survey is an inexpensive and potentially powerful method of identifying and statistically testing patterns across studies (Lucas, 1974). It is particularly suitable when case studies dominate an area of research (Yin & Heald, 1975), when the unit of analysis is the organization, when a broad range of conditions is of interest (Jauch et al., 1980), and when an experimental design is impossible or otherwise fails to capture situations relevant to managerial practice (Bullock & Tubbs, 1987). The basic procedure of the case survey is (1) select a group of existing case studies relevant to the chosen research questions, (2) design a coding scheme for systematic con-

version of the qualitative case descriptions into quantified variables, (3) use multiple raters to code the cases and measure their interrater reliability, and (4) statistically analyze the coded data.

Case Survey Strengths

The four methodological articles cited in the previous paragraph have indicated several strengths of the case survey method. First, it taps prior research efforts reported in a vast number of case studies that contain managerially relevant data due to their reliance on what Argyris (1970, 1980) called organic, action-oriented research (rather than mechanistic, rigorous laboratory experiments). Second, the case survey method overcomes major drawbacks of single case studies, namely, their inability to examine cross-sectional patterns and to generalize to large populations. Pooling relevant case studies into data sets sufficiently large for statistical testing provides the nomothetic advantages of cross-sectional analysis and generalization. Third, the method capitalizes on the idiographic richness of case studies that derives from their ability to study more complex phenomena than more superficial nomothetic surveys can study (cf. Tsoukas, 1989). The typically longitudinal and multisource data collection of case studies captures organizational processes and multiple stakeholder perspectives better (cf. Walton, 1972).

Fourth, case surveys can be replicated since both their coding schemes and case study reports are available to other researchers. Researchers can also apply the coding schemes to other case studies to cross-validate or extend the original findings. Furthermore, the reliability of the case codings can be readily measured through using multiple independent raters to code the same cases and assessing the extent to which they agree. All this eliminates some of the risks associated with the idiosyncratic art of reviewing research, such as ignored information and sample biases (cf. Cook & Leviton, 1980; Glass et al., 1981; Hunter et al., 1982).

Fifth, the case survey method avoids premature exclusion of studies based on a priori judgments about their research designs, publication status, and age, all of which often plague research reviews. Instead, the aim of case surveys is to initially include most studies relevant to the research questions and to test to what extent different research designs, publication features, and time periods studied affect the findings of the case studies. Such broad inclusion enables a researcher to identify the possible effects of those study characteristics by treating them as empirical issues rather than succumbing to exclusion judgments based on prejudice or convenience. If significant differences emerge, it is easier for an investigator to distinguish between the case findings that result from the phenomena studied and those that emanate from the study characteristics and thereby make more informed choices regarding possible exclusion of unduly biased cases. If no significant differences emerge, the researcher can conclude that the findings are robust across study characteristics.

Sixth, the inclusion of case studies from different time periods also

enables the analysis of patterns of complex phenomena over time; for instance, possible effects of organizational learning might be detected. Case studies benefit from longitudinal analysis of certain organizational processes, whereas cross-sectional surveys done at one point in time do not. However, case studies are limited to the longitudinal study of the focal process situated in one time period. Case surveys can be used to compare organizational processes as they occur in different time periods and thus to investigate the impact of different phases of the business cycle, organizational learning over time, and so on. Furthermore, even if a case study spans several business cycles, the singularity of the observed process makes it difficult to distinguish the influence of the business cycle from other influences. Case surveys provide the nomothetic benefit of statistical control of these other influences. It is surprising that only Lucas (1974) pointed out the potential of case surveys to address these time issues that should be central to management inquiry.

Finally, from a broader perspective, the case survey method provides a valuable bridge over other traditional research gaps, such as those between quantitative and qualitative methods and positivistic and humanistic approaches (cf. Lee, 1991). It accomplishes this bridging by quantifying primarily qualitative case studies to enable statistical hypothesis testing of interpretive data. Although several social scientists have advocated the triangulation of quantitative and qualitative methods, substantial integration of fieldwork and survey methods has been rare (Jick, 1979). Case surveys can contribute to this triangulation, which "heightens qualitative methods to their deserved prominence and, at the same time, demonstrates that quantitative methods can and should be utilized in complementary fashion" (Jick, 1979: 610). Similarly, case surveys are a multimethod strategy, allowing a researcher to "attack a research problem with an arsenal of methods that have non-overlapping weaknesses in addition to their complementary strengths" (Brewer & Hunter, 1989: 17).

In sum, the strengths of the case survey method emanate from drawing on the benefits of both the idiographic and nomothetic approaches to overcome their respective drawbacks. A single case study forfeits the opportunity of cross-case pattern analysis that is advantageous for generating theory (Eisenhardt, 1989) as well as theoretical generalization (Yin, 1984). Multiple case studies can achieve this opportunity, but the resource-consuming, intensive research they require typically limits case sets to smaller sizes than are needed to benefit from advanced statistical cross-case analysis (e.g., Eisenhardt & Bourgeois, 1988; Gersick, 1988). Standardization of designs can increase the number of case studies that can be done within a certain resource limitation, but here lies a major crux of the trade-off between idiographic and nomothetic research. Standardization sacrifices the depth of case data and further exposes case studies to such nomothetic weaknesses as common method variance (e.g., Campbell & Fiske, 1959; Podsakoff & Organ, 1986), in which the use of the same researcher and design can create artifactual covariance. The case survey method attempts to cut the Gordian knot

of this crux, not only by extending resource limitations through the use of already existing cases, but also by drawing on the richness of the cases, on different researchers and designs (to overcome the problem of common method variance), and on nomothetic cross-case analysis and statistical generalization.

Case Survey Limitations

Case surveys do, however, suffer from a number of limitations (Bullock & Tubbs, 1987; Yin, 1981; Yin & Heald, 1975). First, the number of available case studies that are relevant to the specific research questions of interest is a limit. Although it is often easy to believe that a sufficiently large number of cases is not available, a surprising number can typically be found through the use of multiple search strategies and sources. Second, Yin (1984) stated that the selection of case studies is beyond the control of a secondary investigator and therefore, case surveys are unlikely to achieve theoretical and statistical generalization. This view ignores, however, the fact that case survey researchers can control the secondary selection of cases to be included in a set. Lucas's (1974) suggestion that analysts utilize sampling parameters and bias analysis actually points out how this control can be accomplished through stratified case selection.

Third, case study reports restrict the information available for case surveys by leaving out much of the collected data because of space limitations. However, by asking case authors to code their own cases, it is possible to access more of their primary data. Fourth, the quality of the case survey is no better than the quality of the case studies it analyzes. Still, both the effects of different research designs and the validity of case codings can be tested, even though the literature has so far given little and fragmented attention to these issues.

Finally, the coding procedure of assigning numbers can unduly simplify the complex phenomena under investigation. Coding simplification is a key issue in case survey methodology since it constitutes the bridge from idiographic richness to nomothetic generality. Although coding necessarily simplifies the information contained in case studies, the central question is rather if the benefits of quantitative analysis across a large number of cases can compensate for this information loss. The ardent idiographic researcher will probably ascribe little value to these benefits and conclude that almost any simplification is unwarranted. Conversely, the equally ardent nomothetic researcher may wonder why one should bother about case studies at all, given their questionable scientific value. Instead of giving all the credit to their own position and no credit to the other, researchers should see the need for a trade-off between the complementary strengths and weaknesses of the two approaches. A researcher designing a study may reasonably expect diminishing marginal utility from extending an already extensive nomothetic survey, by, for instance, adding one more observation to a large sample, or from intensifying an already intensive idiographic study by adding one more variable to a multiaspect case. Diminishing marginal utility favors designs

that balance the nomothetic and idiographic approaches. A key contribution of this study is addressing how a research design can minimize the information loss from coding simplification to further enhance the value of the case survey method.

It should be observed that the qualifications or suggestions for improvement that follow each of the case survey limitations above are not intended to belittle the importance of these weaknesses. On the contrary, as the empirical case survey review below indicates, these limitations have, in one way or another, seriously undermined most of the applications of the method to date. Nonetheless, the limitations seem surmountable through existing but fragmented and little-utilized methodological developments. Bullock and Tubbs (1987), for instance, provided the most comprehensive methodological treatment of the quantitative analysis of case studies, but they referred to only one of the seven prior empirical applications and to none of the three prior methodological articles on such analysis that are reviewed here. Thus, one of the most severe limitations of the case survey method has been remarkably poor use of previous contributions.

EMPIRICAL CASE SURVEY REVIEW

Yin and Yates (1974) claimed their study of decentralization of urban services was the first important test of the case survey approach. They extensively documented a relatively large sample with many variables. Subsequent research has not employed two of their contributions: testing the effects of "research quality" on the case codings and indicating the degree of confidence for each coded variable to distinguish between strong and weak responses. Yin, Heald, and Vogel (1977) further developed the coding of research quality in a similar case survey of technological service innovations, which is also reported in Yin, Bingham, and Heald (1976).

Mintzberg, Raisinghani, and Theoret's (1976) article "The Structure of 'Unstructured' Decision Processes" is probably the best-known report of a case survey. Those authors neither made a methodological point nor used a specific label, but their study fulfills the basic steps of investigating research questions through a coding scheme applied to a set of cases. Although their study lacks interrater reliability, it is nonetheless a fine example of how much even a primitive case survey can accomplish. Miller and Friesen (1977) used a more advanced case survey to isolate different archetypes of strategy making in the context of a host of variables. They contributed the use of seven-point Likert-scale variables and multivariate statistics and had the top executives of the case firms comment on the adequacy of specific case quotes and the general validity of the case. Miller and Friesen (1980a, 1980b) subsequently used a hybrid design combining case survey and primary questionnaire to study organizational adaptation in cases in which the primary data base from firms' executives helped confirm the validity of the case survey codings.

Golembiewski, Proehl, and Sink (1981) investigated the efficacy of organizational development applications in the public sector. Their total group of 574 cases, which includes a comparison group of 304 private-sector applications, constitutes the most extensive case survey reviewed here. But those authors only studied two variables—type of organizational development design and degree of success. Osborn, Jauch, Martin, and Glueck's (1981) study of chief executive officer (CEO) succession, performance, and environmental conditions was the first adequately combined treatment of interrater reliability and discrepancy resolution. Their use of the majority opinion in the disagreement analysis indicated a second important reason to use multiple raters for each case. Previously, multiple raters had only been used on small parts of samples to calculate interrater reliability, but they can also improve the quality of codings through identifying mistakes and personal biases. No rater is infallible, and if extra raters' codings have no impact on a principal rater, the latter's mistakes and personal biases can go undetected. Using the majority code is one way to make the extra codings count to improve the quality of the case data set.

The gainsharing case survey reported in Bullock and Lawler (1985) and Bullock and Tubbs (1987, 1990) is an important milestone in the use of multiple raters. Bullock and Tubbs (1987) listed four techniques for resolving discrepant codings: the use of (1) the codings of one expert rater, (2) the modal score, (3) the average score, and (4) consensus ratings: "All [four] techniques meet reliability standards, but they differ in validity and reproducibility, which we believe are more important than reliability in case meta-analysis. Though average, modal, or single expert ratings are frequently used, we recommend the consensus approach because the data are readily available to improve the potential validity and reliability" (1987: 202–203).

In the consensus approach, multiple raters meet to discuss discrepant codings relative to the case reports studied in order to agree upon the most correct codes. This is a superior way to correct coding mistakes since it eliminates influence from the least correct codings. Coding errors affect average scores, and majority interpretations are not always correct. The consensus approach can also generate additional coding rules that help to resolve discrepant interpretations of unforeseen special cases. In contrast, questionnaires are typically static in that once they are distributed, they cannot be changed in the event of unexpected reactions from respondents, and therefore researchers tend to avoid the negative consequences of such unexpected reactions through simplified design of their questionnaires.

More recent case surveys include Calingo's (1989) study of the environmental determinants of competitive strategies, Tang's (1991) study of the institutional arrangements of common-pool resources, and Manimala's (1992) work on the entrepreneurial heuristics for high- and low-innovation ventures. However, these studies do not provide any significant methodological developments, other than using two of the existing method labels. Tang's study is poorly reported, and the strengths emanating from Calingo's

use of external validation measures and Manimala's extensive efforts with 197 variables and 164 cases are offset by, for example, a narrow search strategy and inappropriate or nonexistent resolution of coding discrepancies.

Case Survey Comparisons

Table 1 compares seven case surveys cited above and a merger and acquisition study that I conducted (Larsson, 1989) in terms of a detailed case survey procedure in which the four major methodological sections are broken down into the following 12 steps: (1) developing initial research questions as a basis for (2) case selection criteria and (3) case sample collection; (4) designing the coding scheme to convert the cases into variables; (5) coding the cases through multiple raters and (6) participating authors, (7) measuring interrater reliability, and (8) resolving coding discrepancies; statistically analyzing (9) the coding validity, (10) the impact of specific case study characteristics, and (11) the created case data set; and finally, (12) reporting the study.

Several patterns emerge from comparing the case survey's handling of the 12 procedural steps. First, the research questions address a diverse set of complex organizational processes. Second, the main selection criteria are what Bullock and Tubbs (1987) called the theoretical domains given by the research questions, that is, the conceptual specification of which empirical phenomena are studied through the case survey. Unfortunately, such domain criteria tend to be complemented with less admissible initial selection criteria regarding publication status and age that can be based on dysfunctional prejudice against unpublished and "dated" findings; application of such criteria is comparable to the bias of journals toward accepting "strong" findings (Rust, Lehmann, & Farley, 1990). The number of cases intended to be collected varied from an unspecified number to all relevant cases or as many as resources allowed. No stratified sampling attempt was made in any of the surveys. Third, although collectively the surveys represented the use of many different search strategies and sources to gather cases, individual case surveys often suffered from the use of few search strategies and sources, both limits that can unnecessarily bias a case data set. Authors made no comparisons with population or other existing sample statistics to analyze possible biases (cf. Lucas, 1974). Table 2 gives a more detailed comparison of the selection criteria and case collection methods used in the case surveys.

Fourth, the coding schemes vary strongly in the numbers of variables used, with a range of from 2 to 197, and in the types of rating scales used (yes-no to seven-point Likert). Thus, there are examples of authors' following two partially contradictory recommendations, one advising use of the simplest coding schemes possible (Bullock & Tubbs, 1987) and the other, development of broad yet precise concepts allowing incorporation of diverse findings (Lucas, 1974).

Fifth, most case surveys have used only one or two raters per case. Sixth,

no case survey reports any author participation in spite of its usefulness for validation purposes (Lucas, 1974) and for the provision of additional information not included in the case reports (Bullock & Tubbs, 1987). Seventh, the interrater reliability measures are difficult to compare not only because they differ, but also because the levels of complexity of the variables in the coding schemes and numbers of raters used differ. For instance, 80 percent absolute agreement between three raters coding many variables with seven-point Likert scales is a much more stringent reliability measure than 80 percent absolute agreement between two raters coding a few yes-no variables; the probability of agreement through random guessing is about 2 percent per variable in the first example, as compared to 50 percent in the second. Approximately two-thirds agreement seems to be viewed as the level of satisfactory reliability (cf. Calingo, 1989; Jauch et al., 1980; Yin & Heald, 1975). Eighth, of the few studies that utilized any form of discrepancy resolution to improve the quality of the codings, only Bullock and colleagues (1985, 1987, 1990) and Tang (1991) used the consensus approach. No analysis of possible resolution biases was made.

Ninth, validity tests have been limited to Miller and Friesen's (1977, 1980a, 1980b) accuracy questionnaire and subsequent hybrid designs, which Manimala (1992) also used, and Calingo's (1989) correlations between the case codings and external measures. Tenth, the almost complete lack of analysis of studies' research designs, publication, and time periods covered is an area of special concern since it undermines the premise of the case survey methodology to initially include all relevant studies. Such broad inclusion requires an analysis of the impact of studies' different characteristics. Lack of these analyses also neglects an excellent opportunity to study such patterns of organizational processes as learning and differences across the business cycle over time. Eleventh, statistical analyses range from simple frequencies to multivariate methods. Twelfth, authors have documented the coding schemes and data sets of their case surveys using books and supplementary methodological articles.

In all, the existing case surveys have much room for improvement. Fully adequate methodological treatment requires that all 12 steps listed above be accomplished—with the possible exception of author participation—since each step provides input essential to the next. Several weaknesses offset the strengths of each study, as the numerous missing or otherwise flawed steps in Tables 1 and 2 indicate. Given that a case survey design is not stronger than its weakest link, it is essential to strengthen the entire case survey procedure.

COMPREHENSIVE DEVELOPMENT OF THE CASE SURVEY METHOD

The development of the case survey method centers on synthesizing fragmented contributions to maximize the amount of information that can be extracted from case studies. I use my recent case survey of mergers and acquisitions (Larsson, 1989), one of the studies included in Tables 1 and 2,

TABLE 1
Methodological Review of Eight Case Surveys[a]

Case Survey Step	Yin & Yates (1974)	Yin et al. (1977)	Mintzberg et al. (1976)	Miller & Friesen (1977)	Golembiewski et al. (1981)	Osborn et al. (1981)	Bullock & Lawler (1985)[b]	Larsson (1989)
1. Research questions	Urban decentralization	Service technology innovations	Unstructured decision processes	Strategy making in context	Public organization development applications	CEO succession, environmental volatility	Gainsharing	Merger and acquisition integration
2. Case selection criteria[c]								
3. Number of cases	215	140	25	81	270 (574)	313	33	55
4. Coding scheme								
Number of variables	118	94	21 (+?)	31	2	19	20	84
Typical scale	Nominal	Nominal	Sequence	Likert-type	Ordinal	Rank 5 of 23 alternative	Nominal	Likert-type
Number of points	2–4	4–5	Open frequency	7	4		2 and insufficient information	5 and insufficient information
Confidence scoring	55% sure	74% sure						
Research design	Internal validity	Additional analysis						Systematic and extent
5. Number of raters per case	1.1	1.2	2	1.4	2	3	3	2.8
6. Author participation			(Teacher)				(Possible development)	7 (26 cases)

TABLE 1 (continued)

Case Survey Step	Yin & Yates (1974)	Yin et al. (1977)	Mintzberg et al. (1976)	Miller & Friesen (1977)	Golembiewski et al. (1981)	Osborn et al. (1981)	Bullock & Lawler (1985)[b]	Larsson (1989)
7. Interrater reliability	82–61%	62% (77–60%)		90% (+/−1)	.78 correlation	90%	(47–93%) (.4–.9 kappa)	**Stepwise average pairwise percent agreement 69 (−50%)**
8. Discrepancy resolution						Majority	Consensus	**Consensus and rater influence analysis**
9. Coding validity tests				Questionnaire on case reports to primary respondents				**Secondary validation by author data**
10. Impact of case characteristics Case collection								
Research design	Yes	Yes						Yes
Publication status		Yes						Yes
Time period		(Yes)						Yes
11. Analysis of data	Bivariate correlation	Bivariate correlation	Frequency count	**Obverse factor analysis**	Frequency count	Multiple discriminant	Bivariate correlation	Multiple regression analysis
12. Reporting study								
Coding scheme	**Including frequencies**	Including frequencies	Described in text	Variables	Described in text	(Other articles)	Examples of text and frequency	Including frequencies
Sample	**Including screened**	Including screened	Table		(Other reference)		(Source type)	Including screened

[a] Boldface type indicates original contributions or specific strengths. Parentheses indicate partial or limited application of the step in question.

[b] Bullock and Tubbs (1987, 1990) also report this case survey.

[c] Table 2 gives further details.

TABLE 2
Review of Case Selection Criteria and Case Collection of Seven Case Surveys[a]

Case Characteristics	Yin & Yates (1974)	Yin et al. (1977)	Mintzberg et al. (1976)	Miller & Friesen (1977)	Osborn et al. (1981)	Bullock & Lawler (1985)[b]	Larsson (1989)
Selection criteria							
Theoretical domain	Urban decentralization	Service technology innovations	Unstructured decision processes	Strategy making	CEO succession, environmental volatility	Gainsharing	Merger and acquisition integration
Further specifications and exclusion rules	Site-specific organizational change	Agencies' first experiences with implemented innovations	Strategic outcome sufficiently complete and detailed description	Data for almost all variables	Sufficiently described environmental strategy and performance	Definition and at least moderately detailed description	Six months or more of integration described; At least two pages each on business and human issues; Test ok
Research design	Posttest exclusion	Test without exclusion					
Publication status	Unrestricted	Not trade journals, etc.	Only nonpublished	Only published	Only published	Unrestricted	Unrestricted
Time period studied	1960–73	1965–75	Around 1970	Unrestricted	1930–74	Unrestricted	Unrestricted
Quantity Stratified sampling	All	All	Most detailed		Resources	All	Resources

TABLE 2 (continued)

Case Characteristics	Yin & Yates (1974)	Yin et al. (1977)	Mintzberg et al. (1976)	Miller & Friesen (1977)	Osborn et al. (1981)	Bullock & Lawler (1985)[b]	Larsson (1989)
Case collection Search strategies	Bibliographies, computer search, several specific journals	Bibliographies, computer search, reference lists in prior literature, several specific journals	Own prior possessions	Case catalogues, *Fortune*	*Fortune*	Own prior possessions, bibliographies, case catalogues, colleagues and experts, several specific journals	Own prior possessions, bibliographies, case catalogues, colleagues and experts, computer search, conference programs, reference lists in prior literature, several specific journals
Sources	Research journals, research books, unpublished papers	Research journals, research books, dissertations and theses, conference papers, unpublished papers, corporate documents	Students	Teaching cases, business literature	*Fortune*	Research books, dissertations and theses, conference papers, business literature, popular press, corporate documents	Research journals, research books, dissertations and theses, conference papers, teaching cases, business literature, popular press, unpublished papers
Collected cases	269 (+ ?)	235	48	81 +	420	35 (+ ?)	105
Final sample	215	140	25	81	313	33	55
Bias analysis							Visual ok

[a] The study by Golembiewski and colleagues (1981) that was included in Table 1 was excluded here because of insufficiently reported information.
[b] Bullock and Tubbs (1987, 1990) also reported this case survey.

to illustrate each of the 12 steps. The limitations of this study are discussed below as empirically generated directions for future methodological development.

Selection of a Relevant Case Data Set

Developing research questions. Research questions are the obvious starting point for a case survey. Contrary to Yin and Heald's (1975) suggestion that questions concerning assessment are more appropriate than the discovery of complex processes, the review clearly indicated that analysis of patterns of complex organizational processes is the comparative strength of the method. Case surveys can be both theory-driven tests of hypotheses and explorations of the contents of a relevant literature. Still, even exploratory case surveys need to be grounded in theory and specific research questions in order for an appropriate group of cases to be selected and for an effective coding scheme to be designed.

The focal merger and acquisition study was an attempt to bridge several gaps in the merger and acquisition literature. The main research question addressed the hypotheses that high synergy realization in mergers and acquisitions requires high potential for synergy, high organizational integration, and low employee resistance. Synergy realization was defined as the actual benefits, such as reduced cost per unit of output, increased income, and so forth, created by the interaction between the joining firms. It is a conceptually advantageous performance measurement for mergers and acquisitions since it avoids the weaknesses of both stock price and accounting measures that either measure anticipated performance (cf. Datta, 1991) or have difficulty discerning between performance resulting from other factors and value capture from value creation resulting from a merger (cf. Haspeslagh & Jemison, 1991). Synergy potential refers to the strategic features potentially allowing corporate combinations to realize synergies, such as relatedness (cf. Chatterjee, 1986; Kusewitt, 1985; Singh & Montgomery, 1987). Organizational integration is the amount of interaction between the joining firms and the amount of coordinative effort they expend to realize synergy potential (cf. Napier, 1989). Finally, the human side of mergers and acquisitions is represented by the extent to which the firms' employees resist the combination and integration processes through exits, absenteeism, shirking, protests, sabotage, and so on (cf. Blake & Mouton, 1985; Buono & Bowditch, 1989; Schweiger & DeNisi, 1991). Although previous research has shown that acquisitions seldom create value for acquiring firms (e.g., Hitt, Hoskisson, Ireland, & Harrison, 1991; Porter, 1987; Ravenscraft & Scherer, 1987), attempts to explain their performance have typically been limited to only one of these issues. Given that mergers and acquisitions may plausibly go wrong as a result of problems concerning any one or combination of these three issues, I included them all in the model to account better for variance in performance and to investigate their relative importance.

Case selection criteria. Criteria should be explicit and based on the theoretical domain the research questions define (Bullock & Tubbs, 1987).

Additional exclusion rules should be stated to delineate which cases belong to a theoretical domain and have at least the minimum amount of reported information about the domain to be meaningful for the case survey. Type of research design, publication status, and time period studied should be used as variables instead of initial selection criteria. The amount of reported data on an empirical occurrence within the theoretical domain should determine if a case is selected rather than how, when, or where it was reported. If there are two case studies of the same empirical occurrence, one should be discarded or both should be combined in one set of codings to avoid counting the occurrence twice (cf. Bullock & Tubbs, 1987).

The question of how many case studies should be collected needs to be addressed. The common ambition of collecting all relevant studies may not only be impossible but also tends to result in dysfunctional time demarcations. Even less ambitious aims can generate more cases than available resources can handle (e.g., Osborn et al., 1981). But statistical hypothesis-testing requires numbers of cases large enough relative to the number of independent variables to be examined (cf. Woodman & Wayne, 1985). Thus, the number of existing cases, available resources, and statistical sufficiency limit quantity. If a sufficiently large group of cases cannot be collected, a researcher can change the research questions, selection criteria, or coding scheme to address these limitations.

Stratified sampling is a potentially resource-consuming solution to the generalization problems raised by Yin (1984). A researcher can control certain characteristics of the cases in a data set by specifying representative distributions from available larger sample or population statistics and randomly drawing from a pool of collected cases according to these distributions (cf. Lucas, 1974). The number of cases with initially underrepresented characteristics will thus limit the final case selection. Consequently, a number of collected and coded cases with initially overrepresented characteristics is then discarded to avoid sample bias and better allow for statistical generalization. In this way, representativeness in several sampling dimensions may be difficult to achieve simultaneously. The replication logic of theoretical generalization can be more easily achieved since the investigator merely has to specify which case characteristics should be varied or held constant and collect cases until those specifications can be met.

The case selection criteria for my merger and acquisition study (Larsson, 1989) were based on the theoretical domain of merger and acquisition integration processes. To be included, case studies had to have at least two pages describing the business side and two pages on the human side of a focal combination, and those descriptions had to cover more than six months of the integration period. The aim was to collect a statistically sufficient data set given both time and financial constraints. Five case characteristics served as variables for comparison with larger sample statistics in the merger and acquisition literature: combination year, combination type (according to the Federal Trade Commission typology), industry, domestic or foreign combination, and subsequent sell-off rate.

Collection of case studies. This process involves several substeps. First, a researcher should identify the cases that he or she is already familiar with before generating the hypotheses to check for possible undue influences on the hypotheses formulation, an issue completely neglected in the case survey literature. Second, as many different search strategies—for example, computer searches, expert consultations, and reference lists searches—as possible should be employed to minimize biases resulting from search strategy characteristics. Third, the search strategies should cover as many sources (e.g., research publications, dissertations, conference papers, teaching cases, business literature, and unpublished sources) as possible to minimize source-specific biases. Fourth, a researcher needs to apply explicit screening procedures to the collected case references and exclude the cases that appear to have too little information or to make for overly high resource demands (cf. Osborn et al., 1981). Finally, initial comparisons between case data set characteristics and existing population or larger sample statistics can indicate biases (cf. Lucas, 1974) and the possible need for stratified sampling. Even if the available resources or cases do not allow for stratified sampling, it is still important to make these comparisons for a more informed discussion of the findings and their limitations.

In my merger and acquisition study, I used multiple search strategies and sources to collect the cases, as Table 2 indicates. The 105 collected case studies were screened into the final group of 55 through an initial reading to determine whether cases fulfilled the inclusion criteria noted above. For example, 14 cases turned out to describe less than six months of the integration period and therefore were excluded. Visual comparison of the case data set with several large merger and acquisition samples indicated similar distributions on the variables of year, industry, combination type, nationality, and sell-off rate. Thus, I made no attempt at stratified sampling.

Design of the Coding Scheme

Coding scheme. The core of a case survey, the coding scheme documents and guides the conversion of qualitative case study data into quantified variables that operationally define the research questions. A main crux of the design of a coding scheme is the basic trade-off between resource-saving, reliable simplicity and information-rich complexity. Even though complex coding schemes with broad ranges of alternative positions for each variable require more work, they capture more information than simple yes-no schemes. However, the more alternatives and the finer the distinctions, the greater the risk that low interrater reliability can undermine the value of the extra work.

The main argument for complex coding schemes is that they enable maximal information extraction since unreliable scales can be collapsed so that reliable distinctions can be made, for instance, a seven-point Likert scale can be collapsed to a five-point scale. In contrast, simple scales cannot be expanded without recoding all the cases since the information contained in the more detailed distinctions is not captured initially. Although high interrater reliability is generally seen as favorable for research measurement

purposes, here it can indicate that more information could have been extracted. Manimala's case survey, for instance, reported very high interrater reliability (.91–.94 correlation), yet he concluded that the study was limited by "its inability to have finer measurements than the 3-point scale" (1992: 491). This alleged inability cannot be known without empirically testing whether or not more detailed scales would yield significantly worse interrater reliability, but no such test was reported.

Bullock and Tubbs (1987: 195) noted the limitations of simple coding schemes, writing that "the pursuit of high reliability could become a dysfunctional end in itself," but still they make the contradictory recommendation that "simpler is better" (1987: 189). Their recommendation maximizes interrater reliability at the expense of additional information that may be extracted with adequate reliability if a more complex scheme is used. In accordance with Ashby's (1960) law of requisite variety, raters need enough coding variety to respond to the rich variety in case studies and thereby avoid unnecessary coding simplification of the idiographic data. Complex schemes give coders the opportunity to prove that their judgment can yield more information than merely mechanical reading. The limit on how much information a case survey can capture is indicated by the point at which interrater reliability becomes unacceptably low because of an overly detailed coding scheme. The only way to determine this level is to start with an overly detailed scheme and gradually simplify it until adequate reliability is reached. In this way, a slightly too complex coding scheme with somewhat low reliability can efficiently achieve maximum information with adequate reliability through minor simplification.

In the focal merger and acquisition study (Larsson, 1989), I designed the coding scheme with intentional complexity to maximize information. Most of the 83 variables had five-point Likert scales with an "insufficient information" alternative indicating when codings could not be made on the basis of the case report. The synergy potential and realization variables were measured in terms of the possible and actual benefits from 11 individually coded synergy sources, including consolidation of purchasing, production, marketing, competition, and R&D (cf. Chatterjee, 1986; Hitt et al., 1991; Lubatkin, 1983). The degree of integration was measured through two variables: (1) the amount of interaction between the joining firms through, for example, restructuring and material flows and (2) the coordinative efforts made to improve the quality of this interaction through special integrators, transition teams, joint planning, and so forth. The resistance of the acquired employees was coded for the first and second halves of the studied integration period. Several variables addressed the research designs, publication, and time frames of the case studies.

Multiple Coding and Interrater Reliability

Case coding. Coding involves first assigning at least two, preferably three, raters to each case. Although using two raters saves resources and reduces the risk of low interrater reliability, using three raters improves the

extraction of information by eliminating mistakes and single-minded interpretations and enabling majority votes to resolve discrepancies. Second, the designer of the coding scheme can participate as a rater to become familiar with the coding process and save resources. Still, most of the raters should be blind to the theoretical hypotheses to safeguard against undue coding influences from the designer. Third, each rater should code at least one pilot case to become familiar with the coding scheme and compare codings for calibration purposes. The use of several pilot cases may improve interrater reliability, but it also depletes scarce resources. Finally, to facilitate discrepancy resolution, raters should note passages in the text that they view as important for their codings as well as make suggestions for possible coding scheme changes.

In my merger and acquisition study, 11 people coded the cases. A senior doctoral student and I coded nearly all the cases individually, and six authors and three extra raters, all except one being experienced merger and acquisition researchers holding doctoral degrees, provided third codings. All raters coded a pilot case and then calibrated their codings by comparing them to a set of predetermined correct codes with extensive comments.

Author participation. Authors can be excellent third raters of their own cases. More complex designs benefit especially from the additional information that the authors of cases can bring to bear on the codings as well as the secondary validation based on their first-hand knowledge of the cases. Idiographic case studies involve many aspects that are subsequently not reported in condensed write-ups. Comprehensive data collection makes case studies more malleable to meta-analyses with somewhat different purposes than the original studies if the unreported data can be accessed through author participation. However, author participation actually invites lower interrater reliability because raters with different amounts of information code the cases. The conflict between information and reliability considerations can be solved by making a distinction between intentional discrepancies introduced to improve information value and discrepancies caused by a lack of clarity in case reports and coding schemes. The subsequent steps of the case survey method address this issue.

In my merger and acquisition study, seven authors, including myself, coded their own cases so that 26 of the 55 cases were coded on the basis of the primary case data. This process was helpful in avoiding numerous insufficient information codings by nonauthor raters. As only ten authors were asked to participate, and the complex coding scheme involved a deterring amount of work, the high rate of interest was encouraging. An eighth author provided additional information without fully coding his cases. The propensity to participate was based on their desire to further use their data, their interest in this new method of going beyond individual case studies, and the possibility of getting comparative data and pattern analyses. Thus, researchers should not forgo soliciting valuable author contributions because of doubts as to whether case authors will be willing to spend time coding their own cases.

Measuring interrater reliability. Interrater reliability is a crucial measure of coding quality, even though pursuing it overzealously has often been shown to have dysfunctional effects. The literature on interrater reliability is diverse, as Jones, Johnson, Butler, and Main (1983) illustrated. Bullock and Tubbs (1987) extensively discussed its application in case surveys and concluded that percent absolute agreement—the simple percentage of cases on which all raters code a variable identically—should be used as the primary reliability index, especially for categorical variables. Whereas they suggested interclass correlation coefficients and Cohen's (1968) kappa as supplementary reliability statistics, I focused on how often contradictory reliability and information considerations can be reconciled by developing the primary percent-agreement index. First, the percent absolute agreement discriminates against the use of three or more raters since the risk that at least one rater will code a variable differently than the others increases with the number of raters. Nor does it capture whether all three (or more) raters disagree with one another or if only one rater disagrees with the others. A researcher can solve this problem by calculating the average pairwise percent agreement (APPA) as the number of pairwise identical codes divided by the total number of pair comparisons.

The average pairwise percent agreement is neutral to number of raters since it captures partial agreements missed by percent absolute agreement. It provides a comparable measure across case surveys with different numbers of raters as well as across cases within the same case survey that for some reason have different numbers of raters. The example of three raters coding a variable 5, 5, and 4 for the same case illustrates the advantage of the average statistic. The percent absolute agreement is 0 percent, but the APPA is 33 percent. If only two of the raters are used, the percent absolute agreement is either 0 percent or 100 percent, depending on which two raters they are. In one out of three situations, higher reliability results merely from reducing the number of raters.

Second, to avoid discriminating against author participation, which intentionally creates discrepancies to provide more information, a second and truer interrater reliability should be computed for only raters with equal information.

Finally, percent absolute agreement discriminates against the use of more detailed scales by not adjusting for the increased probability of chance agreements with fewer categories and ignoring different discrepancy magnitudes for ordinal, interval, and ratio scales. Miller and Friesen (1977) compensated for these problems by computing interrater reliability using similar agreements (those within one point) on their seven-point scales. The consensus resolution approach is a preferable solution because it allows for subsequent collapsing of scales until reliable distinctions can be made. Furthermore, this approach offers the information-enhancing feature of detecting and correcting such coding mistakes as clerical errors, unintentional omissions, and simple misreadings of text and scheme. Normally, coding mistakes are prevented by simplifying a coding scheme, but this solution has

negative information effects. The initial and equal interrater reliability measures reflect neither feature of the consensus resolution approach even though both features substantially improve the quality of the coded data (cf. Bullock & Tubbs, 1990). Thus, a final corrected interrater reliability measure resulting after deduction of identified mistakes and coding scheme changes should be computed to adequately reflect the reliability of the resolved data set that is to be analyzed. The three stepwise average pairwise percent agreement measures provide a more complete account of the whole coding process and do not motivate dysfunctional maximization of reliability at the expense of information extraction.

In my merger and acquisition study, the initial interrater reliability for all variables and all 11 raters was an APPA of 49.6. Elimination of the author raters resulted in an equal information reliability of an APPA of 52.8. Further elimination of identified mistakes through the multirater-consensus resolution design and collapse of the five-point variables that had the most coding discrepancies to three-point scales (these were mostly the 22 synergy source variables) resulted in a corrected interrater reliability of an APPA of 68.8. This level is similar to those of other case surveys of equal complexity that have been considered to have satisfactory interrater reliability. Additional collapses would have only marginally improved the corrected interrater reliability since almost 20 percent of the discrepancies were between a numerical value and an insufficient information coding. Values for the traditional percent absolute agreement were substantially lower, at 37, 41, and 56 percent, solely as a result of this measurement's discrimination of the case survey's use of 2.8 raters per case. If only two raters had been used, the expected value of the percent absolute agreement would have been 68.8 percent, that is, equal to the neutral average pairwise percent agreement measure.

Resolving coding discrepancies. Discrepancy resolution has already been discussed above. Consensus resolution is superior to alternative resolution approaches, but it involves considerable work. Given limited resources, it is advisable to save work by initially discussing general rater differences and coding scheme changes to reduce the number of individual case discrepancies that need to be reexamined. The reexamination process then needs to be analyzed to check the extent to which resolutions have been in favor of (1) the majority or minority, (2) the authors or others, and (3) the designer or others, and (4) to check for other potentially relevant rater differences. Legitimately, authors can be favored when they are in the minority, but substantial favoring of the designer's initial codings against the majority raises the issue of undue influence if it is systematically in the direction of proposed relationships.

In the focal merger and acquisition study, I used a limited version of Bullock and Tubbs's (1987) resource-consuming consensus approach for resolving the around 2,200 data points that suffered from coding discrepancies after the collapse of the unreliable variable scales. I resolved the minor discrepancies with the apparently best solutions indicated by acknowledg-

ing my own mistakes as well as case author and majority codings. The joint reexamination addressed the major discrepancies that needed specific discussion. The resolution process identified 461 mistaken codings, so an average 3 mistakes per coded case could be attributed to the complexity of the coding scheme and a higher propensity for mistakes among junior raters. Largely as a result of their additional information, the author codings clearly dominated the resolution process, with an average of only 4 changed codes per case, as compared to 22 and 31 changes per case for myself and the other raters, respectively. Whereas the author codes were favored against the majority almost 13 times per case, other raters were favored against the majority less than once per case. This indicates little, if any, undue designer influence. The study supported the superiority of Bullock and Tubbs's (1987) consensus approach over both the majority approach, which could not handle the 637 resolutions of the initially tied discrepancies and would have missed 416 resolutions in favor of the minority, thereby losing much of the valuable author participation. The consensus approach was also superior to the average approach, which could have handled the initial ties, but would have instead diluted the highly relevant author codings and let the 461 mistaken codings influence the final decisions.

Statistical Analyses

Coding validity. This analysis can be accomplished in two major ways. First, the use of the secondary data constituted by the case reports can be validated against the authors' primary case data. If the discrepancy resolution strongly favors the authors' initial codes, the final codings of cases with participating authors can be seen as representative of the primary data. If a comparison between these cases and those that lack author participation indicates no systematic differences, it can be said to constitute a secondary validation of the whole case data set. Second, data sources external to a case study can be used for primary validation of the final case codings. Primary respondents in the case organizations, or outside experts, or both can be asked to make their own estimations of a subset of the variables for an agreement or correlation analysis. Miller and Friesen's (1980a, 1980b) hybrid designs provide an example of this primary validation, whereas their 1977 study only validated the case reports and not their codings. Archival accounting and stock market data can be used to validate coded performance variables.

The ideal situation is to have some of a case's main variables entirely measured through data that are independent of the case study to avoid possible problems with common method variance (Kemery & Dunlap, 1986). Such data are often only available for a subsample of the cases because case identities are commonly suppressed and data bases that include all the selected cases are lacking. Still, the case survey method can better safeguard against common method variance than can self-report studies (cf. Podsakoff & Organ, 1986). The typically longitudinal, multisource data collection of case studies is less subject to rationalization after events and social desir-

ability bias than individual self-reports (Miller & Friesen, 1977). Furthermore, multiple raters and consensus resolution make case interpretations a collective rather than a subjective endeavor. Primary and secondary validation procedures are the strongest safeguard since independent primary data and several different researcher purposes, perspectives, frameworks, and methods are brought to bear on the case codings.

In the merger and acquisition study, the dominance of the author codings implies that the resolved data set for the 26 author-coded cases can be seen as representative of primary data collection. The correlation matrix in Table 3 includes a dummy variable for author-coding to test whether those cases systematically differ from the other cases. The overall lack of significant relationships between the variables studied provides secondary validation of the final codings as representative of the primary case data. I did no primary validation of the case codings against independent data.

Impact of specific case characteristics. The analysis of case collection, research design, publication, and time is an empirical investigation of to what extent and how differences in these case study characteristics affect the case codings. The search strategy generating a case should be noted and included as a variable in the data base. The coding scheme should include variables representing (1) aspects of the case study design, such as the extent and types of data collection used, main empirical perspective used, and the validation procedures used; (2) the publication status ranging from, for example, academic journal to unpublished working paper; and (3) the year or years in which the studied phenomena occurred. If significant relationships are found to indicate that the search strategies, research design, or publication status have systematic influence, informed judgments can be made regarding methodological control variables and possible exclusion of cases that suffer from undue influence. Analysts should view systematic variations over time as possible indications of organizational learning or business cycle influences to be investigated, instead of as dated findings to be excluded.

In the merger and acquisition study, I included variables representing how extensive the data collection was, how systematic it was, the length of the integration period studied, the average calendar year of the studied integration period, the main empirical perspective (acquiring firm = 1, mix = 2, acquired firm = 3), and the publication status of the case studies. Correlations between the variables shown in Table 3 indicate that the extent of data collection, the degree of systematic data collection, and the integration year are significantly related to the substantive variables studied. The high correlation (.61) between the two data collection variables indicates that they largely overlapped. When each of the substantive variables was tested as a function of all seven case design control variables, the extent of data collection was only found to be significant when the degree of systematic data collection had been dropped out of the equations. Although this finding could be interpreted to mean that the more data collected, the higher the codings of the studied variables, the opposite direction is more relevant. That is, the higher the synergy potential, integration, and resistance found in

TABLE 3
Means, Standard Deviations, and Pearson Correlation Coefficients

Variables	Means	s.d.	N^a	1	2	3	4	5	6	7	8	9	10	11
Substantive variables														
1. Synergy realization	2.03	1.82	55											
2. Synergy potential	3.91	2.49	55	.63**										
3. Strategic potential	6.11	2.55	53	.53**	.75**									
4. Degree of integration	2.92	1.08	54	.66**	.62**	.57**								
5. Employee resistance	2.58	1.16	48	−.24	.21	.18	.06							
Case design variables														
6. Author participation in coding	0.47	0.50	55	−.07	−.13	−.36**	−.15	−.11						
7. Empirical perspective	1.94	0.65	55	−.07	.01	−.17	−.15	.21	.36**					
8. Extent of data collection	3.44	1.15	50	.35**	.48**	.45**	.38**	.36**	−.19	.41**				
9. Degree of systematic data collection	3.65	1.05	49	.32*	.35**	.34**	.33*	−.02	−.08	−.05	.61**			
10. Length of studied period	3.92	1.30	47	.05	−.18	−.28*	−.19	−.07	.34*	−.08	−.19	−.26		
11. Publication status	3.20	1.01	55	−.21	−.35**	−.47**	−.22	−.27*	.14	.02	−.24*	−.09	.08	
12. Average integration year	3.86	1.03	55	.32**	.24*	.30*	.34**	.03	.17	−.10	.16	.25*	−.21	−.15

[a] Differences are due to insufficient information codings for some variables.

* $p < .05$

** $p < .01$

a given case, the higher the tendency to find and report pertinent merger and acquisition data; unrelated firms that continue to lead largely separate lives after formal combination yielded less pertinent data. Furthermore, the cases in which the amounts of data collected were low were still scholarly and therefore exclusion was not warranted. The significant, negative impact of publication status on synergy potential in particular was largely an artifact resulting from the inclusion of a high-status publication concerning 14 conglomerate acquisitions with very low synergy potential.

Table 4 displays results of the main regression analysis of the substantive variables, in which the extent of data collection and the integration year figure as insignificant control variables (model 1). If the length of the studied integration period was also controlled for (model 2), it and the integration year became significant. One tentative interpretation could be that, once differences between the periods covered by the case studies were controlled for, significant increases in synergy realization appeared over time. Unfortunately, insufficient information codings reduced the number of observations in this regression equation to merely 35.

Analyzing case data. Researchers can carry out the data analysis through conventional statistics, with selection depending on the types of scales used. Although Bullock and Tubbs (1987) recommended using only bivariate statistics for research questions in less developed theoretical areas, some case surveys have shown the usefulness of multivariate statistics. LISREL appears especially relevant for studying the complex relationships typically found in case studies. Podsakoff and Organ (1986) also suggested that analytic package as a means for dealing with common method variance, and Glick, Jenkins, and Gupta (1986) illustrate so using LISREL.

TABLE 4
Results of Regression Analysis with Case Design Variables

| | Synergy Realization | | | | | | | |
| | Model 1 | | Model 2 | | Model 3 | | Model 4 | |
Variables	β	s.e.	β	s.e.	β	s.e.	β	s.e.
Synergy potential	.404**	.100	.378**	.095				
Strategic potential					.329*	.113	.417*	.121
Degree of integration	.305**	.233	.443**	.249	.340*	.249	.440**	.270
Employee resistance	−.403**	.175	−.337**	.178	−.326**	.194	−.279*	.194
Extent of data collection	.137	.316	.162	.290	.125	.221	.096	.205
Average integration year	.148	.247	.250*	.242	.206	.274	.321**	.268
Length of studied period			.349**	.181			.388**	.199
Adjusted R^2	.612		.724		.575		.712	
F	14.23**		15.90**		11.84**		14.18**	
N	43		35		41		33	

* $p < .05$
** $p < .01$

Table 4 also shows the main statistical results for the merger and acquisition study. Findings clearly supported the hypotheses that high synergy realization requires high synergy potential, high integration, and low employee resistance, with those variables explaining more than 60 percent of the variance. Synergy potential is conceptually a more advantageous measurement of the strategic possibilities of corporate combinations than the more commonly used relatedness measures, since the former covers a wider range of possible synergy sources. However, synergy potential is less statistically adequate since by definition it represents the maximum limit of the dependent variable, synergy realization. To estimate the extent to which possible consistency bias inflates the explained variance, I subsequently created a strategic potential variable to replace synergy potential as the strategic independent variable in models 3 and 4, whose results appear in Table 4. Strategic potential was based on previous codings of (1) the degree of similarity between the geographic markets, customer types, product types, production, and industries of the joining firms, which is typically evaluated by relatedness measures of mergers and acquisitions and (2) the strategic complementarity of the joining firms' operations, in terms of vertical economies, new market access, cross-selling, and know-how transfer; these were included to capture synergistic differences not covered by the similarity-oriented relatedness measures. Together with integration and employee resistance, this statistically more adequate variable still explained more than 55 percent of the variance in synergy realization, which suggests little consistency bias.

Even though the small number of cases studied casts doubt on the conclusion that organizations seem to have learned to realize more synergies in mergers and acquisitions over time, other observations also support this temporal finding. Judging from the dominant themes of the latest merger and acquisition waves, organizations seem to have learned the importance of synergy potentials through the errors of the 1960s conglomerate wave; the wave of the 1980s was characterized by related combinations (Reed Lajoux, 1984). The significant, positive correlation of relatedness with integration year in Table 3 provides support, although it may also be partly due to changes in the U.S. antitrust regulations. Furthermore, numerous anecdotal reports in the popular press describing increased top management involvement, specialized task forces, and preplanned integration packages suggest that organizations have also learned that these higher synergy potentials require a high degree of integration to be realized; the significant, positive correlation of degree of integration with integration year is noteworthy. What remains to be learned is how to reduce employee resistance, which shows an insignificant correlation with integration year and is indicated by practitioner uncertainty about cultural, communication, and other human resource issues in mergers and acquisitions (e.g., Buono & Bowditch, 1989; Larsson, 1993; Schweiger & DeNisi, 1991).

Reporting the study. Reporting these 11 steps presents a practical problem, the need for a trade-off between adequate documentation and space

limitations. Authors of previous case surveys have often solved this problem through publishing books and additional methodological articles. Still, the viability of the case survey method depends on the establishment of a condensed reporting system that enables case surveys to be adequately documented in single articles. Such a system could include a divided reference list giving included cases, excluded cases, and other references and an appendix giving the coding scheme. The two case reference lists could utilize general footnotes to indicate the type of search strategy that generated the case in question and which exclusion rules were applied for the excluded cases. The frequency, mean, and standard deviation of each variable can be included in the coding scheme appendix in conjunction with its operational definition. A general note in the text referring readers to the coding scheme for the definitions of variables could prevent having to define them in the text.

The merger and acquisition case survey was thoroughly reported in a book (Larsson, 1989) using divided reference lists and an 18-page appendix giving the coding scheme and descriptive statistics; this appendix is available upon request.

DISCUSSION

Case surveys constitute a powerful method with which to study many issues in some depth across many cases. The method allows nomothetically inclined researchers to add idiographic richness and managerial relevance to their rigorous and generalizable statistical analyses across large data sets. Conversely, the methodology offers researchers with an idiographic orientation the opportunity to systematically compare their own case studies with numerous other case studies and thereby explore wider patterns and test the generality of their unique insights into complex organizational processes. Thus, case surveys enable management research to go beyond the methodological limitations that researchers impose on themselves by strictly adhering to only a nomothetic or idiographic approach. Such adherence has bifurcated management research into two methodological disciplines, each conducting more of the same research with diminishing marginal utility (Larsson & Bengtsson, 1993). The major contribution of this article is the development of the case survey method's capacity to extract reliable information from large numbers of rich case studies and thus facilitate needed cross-fertilization across this methodological gap.

Advantages

The case survey illustration indicates several methodological advantages for the study of merger and acquisition processes. The complex coding scheme allowed for richer data collection and analysis than a questionnaire survey could realistically be expected to achieve. Case surveys also avoid the low response rates that typically plague even simple questionnaires (e.g., Datta, 1991). Synergy realization has conceptual advantages over the com-

mon accounting and stock-price measurements of merger and acquisition performance, but it is much harder to objectively measure than the other two. I overcame these measurement obstacles through the case survey method's ability to capture rich information. Even if the different case reports tended to focus on either the business or the human side of mergers and acquisitions, the comprehensive data collection inherent in idiographic case studies allowed for an analysis of the broader research question of how these two sides of mergers and acquisitions came together in the organizational integration process. In this way, the case survey method contributes to bridging the fragmented merger and acquisition research.

The empirical application illustrates several of the benefits of the developed comprehensive case survey methodology. I found the average pairwise percent agreement measure of interrater reliability to be superior to percent absolute agreement because it prevented dysfunctional discrimination against the information-enhancing use of more than two raters. Bullock and Tubbs's (1987) consensus approach was also shown to be better than the alternative approaches to resolving coding discrepancies. This case survey further illustrated the information benefits of an initially complex coding scheme as well as the information losses from reliability maximization. The 22 original five-point Likert-scaled synergy items were collapsed into three-point scales. If they had been further collapsed into simple binary items, the interrater reliability would have increased from 68.8 percent to about 71 percent. As an indication of the information loss incurred with this simplification, it is notable that the amount of variance explained by the three main independent variables would have then been less than 50 percent, as compared to more than 60 percent when the three-point scales were used. An initially simpler coding scheme with only binary synergy items would have achieved slightly higher interrater reliability, but I would then never have been aware that more than 60 percent of the variance could be explained through a more complex coding scheme, with a minimal loss of interrater reliability.

A major conclusion is that coding discrepancies should be embraced rather than avoided through initial simplification and use of few raters. Coding discrepancies are the key to identifying the degree to which information can be extracted before it becomes too unreliable, mistakes and less relevant interpretations that otherwise would not have been reexamined, and additional information provided by author participation. The intentional creation of these desirable discrepancies through complex coding schemes and use of more than two raters (including authors) results in lower traditional measures of interrater reliability. The introduction of the stepwise APPA measurement enables the use of these information-enhancing designs without being compromised by lower interrater reliability scores.

The case survey method's unique combination of idiographic and nomothetic strengths makes it very suitable for broad studies of complex processes that have been often researched through case studies, such as sociopsychological micro processes and strategic and organizational macro pro-

cesses, and it is particularly useful for studying processes spanning both levels. Case surveys are also particularly valuable for researching the development of complex patterns over time. The apparent pattern that emerged in the focal case survey, in which organizations learned to acquire higher synergy potential and use a higher degree of integration, could not have been seen through a cross-sectional survey or a small number of case studies.

The possibility of both exploring "what the existing literature says" (Yin, 1984: 118) and testing hypotheses through case surveys could make the method central to the overall interactive research process of examining data, theorizing, testing, reexamining data, and so on. Researchers could also combine the case survey method with more traditional methods in the attempt to bridge the nomothetic-idiographic gap by going back and forth between these polar approaches (cf. Lee, 1991; Luthans & Davis, 1982). For example, case studies could be used to generate an initial in-depth understanding of a small number of empirical occurrences. The set of primary case studies could then be supplemented with existing case studies to create a case survey data set to extend and modify the initial case findings. Once the resulting improved understanding had been crystalized into parsimonious theory, it could be nomothetically tested through a large-sample survey. Finally, researchers could also easily further explore new questions generated by the interpretation of case survey tests by returning to the available and now familiar rich case studies and perhaps even by designing a supplementary coding scheme for relevant issues not covered in the first attempt.

Implications

Some weaknesses in the merger and acquisition application provide directions for future methodological development. First, in attempting to overcome the lack of primary validation, I faced difficulty collecting independent data on the case studies because the identities of firms were often suppressed and the firms were dispersed over more than ten countries. Securing such data could be addressed as early as the case selection and collection phases. Second, the use of only multiple regression equations limited the statistical analysis of the complex relationships among the more than 20 substantive and design control variables. The suitability of case surveys for studying complex organizational processes and the need to control for several study characteristics as well as for common method variance clearly suggest that LISREL analysis is likely to be a more adequate statistical method.

Third, possible problems with distinguishing between different numerical values on overly detailed scales can be handled through additional anchoring and subsequent collapses. The main limitation of the corrected interrater reliability in this study was the difficult distinction between a numerical value and insufficient information (cf. Yin et al., 1977). Further methodological development is needed for the design of coding schemes and resolution processes that can better deal with these common discrepancies. Fourth, more sophisticated variables, based on different validation tech-

niques, maximization or minimization of similarities between selected comparison groups (cf. Glaser & Strauss, 1967), and so forth, are needed to account for methodological differences in the data collection and analysis of case studies. Case surveys actually offer excellent opportunities to review the impact of different case study methods. Besides controlling for their possibly moderating effects on the substantive research questions, case survey analyses can comparatively evaluate and thereby systematically enhance case study methodologies.

Finally, both idiographic and nomothetic researchers can be expected to be wary of the interpretations of interpretations involved in the case survey method. Yin's (1984) later preference for multicase studies by the same researcher over case surveys similarly favors closeness to the primary data. Still, multicase studies are weaker than case surveys from the viewpoint of common method variance since multicase studies are performed by the same person with the same purpose, perspective, method, and theoretical framework. The case survey methodology presented here offers a stronger test by attempting to find valid and reliable patterns across diverse sets of rich case studies conducted by several authors with varying purposes, perspectives, methods, and theoretical frameworks. It provides systematic, replicable, and intersubjective interpretation of a primary researcher's interpretations and thereby controls for his or her idiosyncrasies. Such researcher idiosyncrasies can be a barrier as significant to the generalization of case studies as the use of a small number of cases.

The case survey methodology offers a complementary way of nomothetically using idiographic findings without either discarding the benefits of one of the two approaches or suffering from their weaknesses. The method's ability to study many issues deeply across large numbers of cases can empower researchers to cross new frontiers in management research.

REFERENCES

Argyris, C. 1970. *Intervention theory and method: A behavioral science view.* Reading, MA: Addison-Wesley.

Argyris, C. 1980. *The inner contradictions of rigorous research.* New York: Academic Press.

Ashby, W. R. 1960. *Design for a brain.* London: Chapman & Hall.

Bangert-Drowns, R. L. 1986. Review of developments in meta-analytic method. *Psychological Bulletin,* 99: 338–399.

Blake, R. R., & Mouton, J. S. 1985. How to achieve integration on the human side of the merger. *Organizational Dynamics,* 13(3): 41–56.

Brewer, I., & Hunter, A. 1989. *Multimethod research: A synthesis of styles.* Newbury Park, CA: Sage.

Bullock, R. J., & Lawler, E. E. 1985. Gainsharing: A few questions, and fewer answers. *Human Resource Management,* 23: 23–40.

Bullock, R J., & Tubbs, M. E. 1987. The case meta-analysis for OD. *Research in organizational change and development,* vol. 1: 171–228. Greenwich, CT: JAI Press.

Bullock, R. J., & Tubbs, M E. 1990. A case meta-analysis of gainsharing plans as organization development interventions. *Journal of Applied Behavioral Science,* 26: 383–404.

Buono, A. F., & Bowditch, J. L. 1989. *The human side of mergers and acquisitions.* San Francisco: Jossey-Bass.

Burrell, G., & Morgan, G. 1979. *Sociological paradigms and organizational analysis.* London: Heinemann.

Calingo, L. M. R. 1989. Environmental determinants of generic competitive strategies: Preliminary evidence from structured content analysis of *Fortune* and *Business Week* articles (1983–1984). *Human Relations,* 42: 353–369.

Campbell, D. T., & Fiske, D. W. 1959. Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin,* 56: 81–105.

Chatterjee, S. 1986. Types of synergy and economic value: The impact of acquisitions on merging and rival firms. *Strategic Management Journal,* 7: 119–139.

Cohen, I. 1968. Weighted kappa: Nominal scale agreement with provision for scaled disagreement or partial credit. *Psychological Bulletin,* 70: 213–220.

Cook, T. D., & Leviton, L. C. 1980. Reviewing the literature: A comparison of traditional methods with meta-analysis. *Journal of Personality,* 48: 449–472.

Cooper, H. M. 1984. *The integrative research review: A systematic approach.* Beverly Hills, CA: Sage.

Datta, D. K. 1991. Organizational fit and acquisition performance: Effects of post-acquisition integration. *Strategic Management Journal,* 12: 281–297.

Eisenhardt, K. M. 1989. Building theories from case study research. *Academy of Management Review,* 14: 532–550.

Eisenhardt, K. M., & Bourgeois, L. J. 1988. Politics of strategic decision making in high-velocity environments: Toward a midrange theory. *Academy of Management Journal,* 31: 737–770.

Gersick, C. 1988. Time and transition in work teams: Toward a new model of group development. *Academy of Management Journal,* 31: 9–41.

Glaser, B. G., & Strauss, A. L. 1967. *The discovery of grounded theory: Strategies for qualitative research.* Chicago: Aldine.

Glass, G. V., McGaw, B., & Smith, M. L. 1981. *Meta-analysis in social research.* Beverly Hills, CA: Sage.

Glick, W. H., Jenkins, G. D., & Gupta, N. 1986. Method versus substance: How strong are underlying relationships between job characteristics and attitudinal outcomes? *Academy of Management Journal,* 29: 441–464.

Golembiewski, R. T., Proehl, C. W., & Sink, D. 1981. Success of OD applications in the public sector: Toting up the score for a decade, more or less. *Public Administration Review,* 41: 679–682.

Haspeslagh, P. C., & Jemison, D. B. 1991. *Managing acquisitions: Creating value through corporate renewal.* New York: Free Press.

Hitt, M. A., Hoskisson, R. E., Ireland, R. D., & Harrison, J. S. 1991. Effects of acquisitions on R&D inputs and outputs. *Academy of Management Journal,* 34: 693–706.

Hunter, J. H., Schmidt, F. L, & Jackson, G. B. 1982. *Meta-analysis: Cumulating research findings across studies.* Beverly Hills, CA: Sage.

Hyman, H. M. 1972. *Secondary analysis of sample surveys.* Middletown, CT: Wesleyan University Press.

Jauch, L. R., Osborn, R. N., & Martin, T. N. 1980. Structured content analysis of cases: A complementary method for organizational research. *Academy of Management Review,* 5: 517–525.

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

Jones, A. P., Johnson, L. A., Butler, M. C., & Main, D. S. 1983. Apples and oranges: An empirical comparison of commonly used indices of interrater agreement. *Academy of Management Journal,* 26: 507–519.

Kemery, E. R., & Dunlap, W. P. 1986. Partialling factor scores does not control method variance: A reply to Podsakoff and Todor. *Journal of Management,* 12: 525–530.

Kiecolt, K. J., & Nathan, L. E. 1985. *Secondary analysis of survey data.* Beverly Hills, CA: Sage.

Kusewitt, J. B. 1985. An exploratory study of acquisition factors relating to performance. *Strategic Management Journal,* 6: 151–169.

Larsson, R. 1989. *Organizational integration of mergers and acquisitions: A case survey of realization of synergy potentials.* Lund, Sweden: Lund University Press.

Larsson, R. 1993. Barriers to acculturation in mergers and acquisitions: Strategic human resource implications. *Journal of European Business Education,* 2(2): 1–18.

Larsson, R., & Bengtsson, L. 1993. *The case survey method: Extending Nordic case study research.* Paper presented at the Nordic Business Administration Conference, Lund, Sweden.

Lazarus, R. S. 1971. *Personality* (2nd ed.). Englewood Cliffs, NJ: Prentice-Hall.

Lee, A. S. 1991. Integrating positivist and interpretive approaches to organizational research. *Organization Science,* 2: 342–365.

Lubatkin, M. 1983. Mergers and the performance of the acquiring firm. *Academy of Management Review,* 8: 218–225.

Lucas, W. A. 1974. *The case survey method: Aggregating case experience,* R-1515-RC. Santa Monica, CA: Rand Corporation.

Luthans, F., & Davis, T. R. V. 1982. An idiographic approach to organizational behavior research: The use of single case experimental designs and direct measures. *Academy of Management Review,* 7: 380–391.

Manimala, M. J. 1992. Entrepreneurial heuristics: A comparison between high PI (pioneering-innovative) and low PI ventures. *Journal of Business Venturing,* 7: 477–504.

Miller, D., & Friesen, P. H. 1977. Strategy-making in context: Ten empirical archetypes. *Journal of Management Studies,* 14: 253–280.

Miller, D., & Friesen, P. H. 1980a. Archetypes of organizational transition. *Administrative Science Quarterly,* 25: 268–292.

Miller, D., & Friesen, P. H. 1980b. Momentum and revolution in organizational adaptation. *Academy of Management Journal,* 23: 591–614.

Mintzberg, H., Raisinghani, D., & Theoret, A. 1976. The structure of "unstructured" decision processes. *Administrative Science Quarterly,* 21: 246–275.

Morey, N. C., & Luthans, F. 1984. An emic perspective and ethnoscience methods for organizational research. *Academy of Management Review,* 9: 27–36.

Napier, N. K. 1989. Mergers and acquisitions. Human resource issues and outcomes: A review and suggested typology. *Journal of Management Studies,* 26: 271–289.

Osborn, R. N., Jauch, L. R., Martin, T. N., & Glueck, W. F. 1981. The event of CEO succession, performance, and environmental conditions. *Academy of Management Journal,* 24: 183–191.

Podsakoff, P. M., & Organ, D. W. 1986. Self-reports in organizational research: Problems and prospects. *Journal of Management,* 12: 531–544.

Porter, M. E. 1987. From competitive advantage to corporate strategy. *Harvard Business Review*, 65(3): 43–59.

Ravenschaft, D. J., & Scherer, F. M. 1987. *Mergers, sells-off, & economic efficiency.* Washington, DC: Brookings Institution.

Reed Lajoux, A. 1984. Mergers: The last frontier for American business. *New Management*, 1(2): 22–29.

Rust, R. T., Lehmann, D. R., & Farley, J. U. 1990. Estimating publication bias in meta-analysis. *Journal of Marketing Research*, 27: 220–226.

Schweiger, D. M., & DeNisi, A. S. 1991. Communication with employees following a merger: A longitudinal field experiment. *Academy of Management Journal*, 34: 110–135.

Singh, H., & Montgomery, C. A. 1987. Corporate acquisition strategies and economic performance. *Strategic Management Journal*, 8: 377–386.

Tang, S. 1991. Institutional arrangements and the management of common-pool resources. *Public Administration Review*, 51(1): 42–51.

Tsoukas, H. 1989. The validity of idiographic research explanations. *Academy of Management Review*, 14: 551–561.

Walton, R. E. 1972. Advantages and attributes of the case study. *Journal of Applied Behavioral Science*, 8(1): 73–78.

Woodman, R. W., & Wayne, S. J. 1985. An investigation of positive-findings bias in evaluation of organization development interventions. *Academy of Management Journal*, 28: 889–913.

Yin, R. K. 1981. The case study crisis: Some answers. *Administrative Science Quarterly*, 26: 58–65.

Yin, R. K. 1984. *Case study research: Design and methods.* Beverly Hills, CA: Sage.

Yin, R. K., Bingham, E., & Heald, K. A. 1976. The difference that quality makes: The case of literature reviews. *Sociological Methods and Research*, 5: 139–156.

Yin, R. K., & Heald, K. A. 1975. Using the case survey method to analyze policy studies. *Administrative Science Quarterly*, 20: 371–381.

Yin, R. K., Heald, K. A., & Vogel, M. E. 1977. *Tinkering with the system.* Lexington, MA: Lexington Books.

Yin, R. K., & Yates, D. 1974. *Street-level governments: Assessing decentralization and urban services,* R-1527-NSF. Santa Monica, CA: Rand Corporation.

Rikard Larsson received his Ph.D. degree in business administration at University of Southern California and his D.B.A. degree at Lund University in Sweden. He is an international research scientist and assistant professor at Department of Business Administration, School of Economics and Management, Lund University. His research interests include various intersections of the strategy, organization, and human resource management areas, such as mergers and acquisitions and other types of strategic changes, service, and coordination frameworks, as well as development of synergistic combinations of different research methods.