Secondary Research

Evaluating Secondary Sources

Contributors: David W. Stewart & Michael A. Kamins
Editors: David W. Stewart & Michael A. Kamins
Book Title: Secondary Research
Chapter Title: "Evaluating Secondary Sources"
Pub. Date: 1993
Access Date: October 11, 2013
Publishing Company: SAGE Publications, Inc.
City: Thousand Oaks
Print ISBN: 9780803950375
Online ISBN: 9781412985802
DOI: http://dx.doi.org/10.4135/9781412985802.n2
Print pages: 17-33

http://dx.doi.org/10.4135/9781412985802.n2

Evaluating Secondary Sources

[p. 17 ↓]

Not all information obtained from secondary sources is equally reliable or valid. Information must be evaluated carefully and weighted according to its recency and credibility. When evaluating secondary information, six questions must be answered: (1) What was the purpose of the study? (2) Who collected the information? (3) What information was actually collected? (4) When was the information collected? (5) How was the information obtained? (6) How consistent is the information with other sources?

The regular user of secondary information often develops a healthy skepticism about information provided by others. There are many ways that data may be misleading if they are not evaluated carefully. Data collection is usually purposive, and the purpose for which information is obtained and analyzed may influence the conclusions drawn, the data collection procedure employed, the definitions of terms and categories, and even the quality of the information. In addition, not all secondary sources, even those that appear on the surface to be relevant, are necessarily appropriate for a given purpose or analysis. A particular source may have information that is similar to what is needed, but may provide measures that use a different unit of analysis from that of interest, focus on a slightly different issue, or otherwise fail to provide the type of information sought.

As an example, consider the case of *Tambrands vs. the Warner-Lambert Company*, makers of the home pregnancy test EPT Plus. Based upon a research study, Warner-Lambert made the advertised claim that their test gave results in “as soon as 10 minutes.” Tambrands, a competitor, found this claim to be exaggerated and brought suit appropriately questioning the validity of the research underlying the claim. In their defense, Warner-Lambert reported the results of a research study of 19 pregnant women who were selected to take the test. The findings revealed that 10, or 52.6%, obtained positive (i.e., pregnant) test results within the claimed 10 minutes; however, 2 women received these results within 30 minutes, and the remaining 7 recorded false

negatives. The argument presented by Warner-Lambert was that their advertised claim **[p. 18 ↓]** was supported by the research findings because "the *overwhelming majority* of women … will in fact obtain accurate results in ten minutes, even on Day 1" (emphasis added). Tambrands countered by appropriately examining how the information was obtained and if reasonable conclusions were drawn from it.

The answer to the first question revealed that the 19 women sampled were actually enrolled at a Cincinnati fertility clinic, so the sample was not representative. Moreover, the claimed "overwhelming majority" of 52.6% was not revealed to be statistically significant from one half ($t = .23$, $p < .95$). Indeed, for such a result to be statistically significant from 50%, a sample of approximately 1,400 pregnant women would have had to be taken. Although the court recognized the questionable statistical validity of Warner-Lambert's data, it ruled against the company on the grounds that "the results … do not support the defendant's claims." If the survey results had been more supportive of Warner-Lambert, it is not clear that issues related to the validity of the study would have come to the forefront. *Thus the lesson to be learned is always to question the information collected and reported by others, as Tambrands did.* No data should be used without careful evaluation, and data obtained from secondary sources require especially close scrutiny.

The evaluation of secondary data should follow the same procedures employed in the evaluation of primary data. The researcher who uses secondary sources, however, does have an advantage. Because the information already exists in some form, evaluation of the quality and appropriateness of the information can be done well in advance of its actual use. Secondary data can be identified and evaluated in a stepwise fashion. Too often researchers wait until there is an immediate need for information before evaluating the appropriateness of existing sources. This leaves little time for careful evaluation and frequently affords no opportunity for searching for more appropriate sources. It is important for researchers who use secondary sources to identify them early and to clarify as much information as possible before beginning analysis.

Questions concerning the source(s) of the data (as above), the measures used, the time of data collection, and the appropriateness of analyses and conclusions should be

raised routinely. The questions a user of secondary sources might raise can be grouped into six broad categories:

It is impossible to evaluate information without knowing the answers to each of these questions. One should be immediately suspicious of any information for which answers to these questions are unavailable. The importance of each of these questions is discussed in further detail in the remainder of this chapter.

WHAT WAS THE PURPOSE OF THE STUDY?

Information rarely is collected without some intent. The intent of a particular study may significantly influence the findings. Data collected to further the interests of a particular group or organization are especially suspect, as the example above suggested. The degree of precision, the types of categories used, and the method by which data are collected and reported are often dictated by the intent of the study.

Consider the following: To support their brand in a comparison against Winston Lights, Loews Theaters (makers of Triumph cigarettes) conducted a four-question consumer survey designed to examine the quality of its brand versus that of Winston. On two questions that measured “preference” and “better taste,” Triumph fared better than its competition. On scales relating to “amount of taste” and “satisfying quality,” however, this was not the case. In spite of this, Loews chose to base their advertising claim on the favorable results to the first two questions only, largely ignoring the remaining questions. This resulted in a lawsuit filed by R. J. Reynolds (manufacturer of Winston Lights) against Loews Theaters. The court ruled in favor of R. J. Reynolds because Triumph had “failed to establish a basis” on which to claim that only the first two questions had relevance to the issue of quality. It also contended **[p. 20 ↓]** that “failure to disclose a material aspect of the results, relating to taste, under the circumstances is misleading.”

Thus, in evaluating secondary research, one must always ask whether the purpose of the study was to reach a preestablished conclusion. The researcher should then be

aware of techniques used (e.g., reporting “cherry-picked” empirical results, as above) to arrive at such preordained results.

Even when the data are not collected for purposes of advocating a particular position, the purpose of the study may confound the interpretation of the data. For example, the best-known measure of price movements in the United States is the Consumer Price Index (CPI) calculated monthly by the U.S. Bureau of Labor Statistics. This index is based on the prices of about 400 items of consumption. The price of each item contributing to the index is calculated by surveying wage earners and clerical workers in some base year and computing the average price paid for each item. The index represents an average for a family of four (father, age 38; nonworking mother; boy, age 13; and girl, age 8) living in an urban area. Thus the index is not representative of the expenditures of most families. It is only a very rough index of what is happening to purchasing power and is not often useful for specific decisions where a high degree of precision is required or where expenditure patterns are different from those used to define the index.

WHO WAS RESPONSIBLE FOR COLLECTING THE INFORMATION?

Information from certain sources may be more credible than information from others. This arises not just from the biases that may be at work, but also from differences in technical competence, resources, and quality. Some organizations have developed reputations for excellent quality control work and for the integrity of their data. Others have reputations for poor work. Generally, those sources of high integrity will provide sufficient information about how the information was obtained to enable a review of the technical adequacy of the data. Learning about the reputations of various sources of information requires investigating their previous work. Contacting clients and others who have used information supplied by the organization will also provide some indication of the reputation of an organization. One might also examine the training and expertise present in an organization supplying information.

[p. 21 ↓]

It is also worthwhile to determine whether the organization that sponsored or conducted the research had a vested interest in any particular outcome. For example, an organization that reports a study of its own effectiveness might have a vested interest in accentuating the positive. A rather sizable industry exists to produce what is often called “advocacy research.” Such research is not designed to produce unbiased answers to questions. Rather, the research is conducted for the purpose of providing support for a particular conclusion or position. Although such research may still yield insights, it must be interpreted with caution.

WHAT INFORMATION WAS ACTUALLY COLLECTED?

In the early 1950s, a congressional committee published an estimate of the annual “take” from gambling in the United States. The figure, $20 billion, actually was picked at random. One committee member was quoted as saying, “We had no real idea of the money spent. The California Crime Commission said $14 billion. Virgil Peterson of Chicago said $30 billion. We picked $20 billion as the balance of the two” (Singer, 1971, p. 410). Here is an example of information entered into the public record that had no empirical basis. No data were collected at all; only a couple of opinions were sought and averaged. “Mythical numbers,” as Singer (1971) refers to them, are more common than one would wish. These mythical numbers, estimates based on pure guesswork, represent the extreme case, but they serve to emphasize the need for asking what information actually was collected.

Consider the following claim by an advertiser: “Mothers say that Brand X diapers absorb 17.5% more liquid than Brand Y.” The first point to be made here is that probably no mother ever said this. Most probably, the figures are derived from a marketing research study commissioned by Brand X and averaged to arrive at this conclusion. The next point to be made is that no matter how accurately this proportion may summarize the results obtained, the data were derived from a sample of mothers. Hence sampling variation is always at issue and may be quite large if proper statistical sampling techniques were not observed. A third issue is that the statement only mentions

mothers. In this day and age, fathers have been known to change a diaper or two … so did their opinions not count?

Taking these issues into consideration, we still must ask, what is the meaning of the advertised statement? Specifically, if Brand Y only **[p. 22 ↓]** absorbs 1 ounce of liquid, then how valuable is a diaper that can only absorb 1.175 ounces? Hence all of these criticisms leave the relevance of such an exact figure as 17.5% in doubt.

The context in which data are collected may also influence the results. Consider a study of consumer preferences that found that 60% of all consumers preferred Brand A. Such a finding is impressive until one learns that brands B and C, the major competitors of A, were not included on the list from which consumers were to select a product.

Many of the things we wish to measure cannot be observed directly. Thus we obtain an estimate indirectly by using a surrogate measure that is observable and assumed to be related to the more interesting phenomenon. The critical assumption of such indirect measurement techniques is that there is a relationship between the observable measure and the unobservable event of interest. Even when this assumption is correct, however, the relationship may be decidedly less than perfect. Consider studies of the success of graduates of corporate training programs. Success is difficult to measure because it involves a variety of dimensions and could be measured at many different points in time. One organization may report results using turnover during the year following completion of the training program. A second organization may use rapidity of advancement within the organization and salary increases during a 3-year period. Still another organization may use ratings of success by supervisors after 6 months on the job. In each case, the data may be used to relate completion of the training program to success on the job, yet the relationship reported may vary widely from one study to another. The differences in the findings are attributable to the data that actually were collected, not what these data were interpreted to mean. Knowing what information actually was obtained is often very useful for reconciling conflicting results. For example, it is well known that self-report data about behavior differ significantly from data about the incidence of the same behavior obtained by observation (Fiske, 1971).

Even when direct measurement is possible, the ways in which data are defined and classified may confound the interpretations made. Categorizations and classifications

may vary widely, and their relevance and meaning for a particular purpose must always be investigated. For example, what is a family? Is a single, self-supporting person living alone a family? Are unmarried cohabitants a family? For some purposes and in some studies the answer is likely to be yes, whereas in other cases the answer is likely to be no. Wasson and Shreve (1976) provide an example of the problems caused by insufficient attention to the classification **[p. 23 ↓]** issue. For many years, the steel industry used total tonnage sold as a criterion of success. The criterion led the industry to overlook its losses of highly profitable low-tonnage sales to paper and aluminum products. Only too late did the industry recognize that a classification system based on uses and markets would have provided greater insight into events in its marketplace.

Wide variations in geographic, income, and age groupings across studies are quite common. There are often no accepted definitions for the concepts measured. Thus careful attention must be given to what information actually was obtained in a particular study. Apparent inconsistencies across studies often have more to do with the operational definition of terms than the actual differences in the underlying phenomena. Of course, such problems hinder the effective usage of meta-analysis across studies and the effective generalizability of conclusions.

The quality and relevance of information may also vary within a given source. For example, a particular data base of mental health hospitalizations was constructed for the purpose of facilitating reimbursements of expenses by such third parties as insurance companies and government agencies. Information on the length of hospitalization tended to be quite good, because this was directly relevant to the issue of reimbursement. Other information within the data base, however, was of more questionable quality. Data on discharge status were particularly poor, because they were never verified. Such data were not of central importance to the original purpose of the data base. The lesson in this example is that a source of information may be very useful for one purpose, but very poor for other purposes. Sources are not good or bad; they are useful for some purposes and not for others. The same is true for individual variables within a given source.

WHEN WAS THE INFORMATION COLLECTED?

A study of the perception of the price of long-distance telephone calls found that consumers were very much aware of the price of long-distance calls and very sensitive to even small rate hikes. The results of the study might be interpreted as an indication that consumers are very price sensitive. The study, however, was carried out while an intense, highly publicized debate over a telephone price hike raged, a debate that included several prominent politicians involved in an election campaign. It is likely that the results of the study would have been different **[p. 24 ↓]** had the study been carried out when there was less publicity about telephone rates.

Time is an important factor to be considered when evaluating information. As in the example above, factors present at the time of information collection may influence the results obtained. Time may also influence the definition of measures. For example, when is a sale made? Does the sale occur upon the placement of an order, receipt of the order, the time of shipment, the time of delivery, the date of billing, the date of payment, or the date a payment actually is recorded? Different accounting systems place emphasis on different points in time and produce differences in information. Shifts in the point of time when measurements are taken may have very pronounced effects on the results obtained.

The passage of time may also change the measurement instrument. Consider the following example provided by Wasson and Shreve (1976). In most places, the dividing line between petty and grand larceny was $50.00 in previous generations and is now $100.00. In 1910, $50.00 represented 2 months' wages, whereas today it may represent less than a day's wages. Thus it may appear that the level of crime has increased when in fact it may have been decreasing.

Time may also make information obsolete. Data on unemployment rates in the 1960s are not particularly useful for formulating policy in the 1990s. Technological changes may change perceptions; life-styles may change. Sooner or later, most secondary data become obsolete and of interest only for historical purposes. How quickly data become

obsolete depends on the type of data, the purpose for which they are used, and what new data have been obtained. In the case of census data, it typically takes at least 2 years before they are published. By nature, its value quickly diminishes over time. Although periodic updates are offered by the Census Bureau itself, as well as local planning and commercial agencies, the updates generally apply to large geographic areas, so primary research may be needed to obtain data on local areas.

The user should always know when data were collected, however, particularly because there is often a substantial time lag between data collection and the publication of results. Some data remain valid despite the passage of time, of course. For example, studies of verbal learning carried out in the 1880s remain useful even today. More recent research has added to our understanding of the learning process, however, and some conclusions have been modified as new information has been obtained. For more information on the temporal boundaries of data, the interested reader should consult Kelly and McGrath (1988) in this series.

[p. 25 ↓]

HOW WAS THE INFORMATION OBTAINED?

The quality of secondary data cannot be evaluated without knowledge of the methodology employed when collecting the data. Information about the size and nature of samples, response rates, experimental procedures, validation efforts, questionnaires, interview guides or protocols, and analytic methods should be available in sufficient detail to allow a knowledgeable critique of the data collection procedure. The following examples help illustrate why such information is useful.

Consider a poll that finds that 80% of the respondents in a survey opposed gun control. One's interpretation of the 80% figure would be quite different if one were to learn that the respondents were drawn from the membership roster of the National Rifle Association and not a representative sample of the total population. Suppose that a report on road accidents in Country X claimed that motorcyclists suffered a great

increase in casualties during the year, based on the findings that there were 1,280 fatalities in 1990 and 1,586 in 1991 (an increase of 23.91% over the previous year). Although the data would seem to support the claim, suppose that a more penetrating examination of available information revealed that the number of registered cycles increased from 1,460,000 in 1990 to 1,910,000 in 1991 (an increase of 30.82%). Hence the rate of accidents per registered cycle, which was 0.088% in 1990, fell to 0.083% in 1991. Examination of these more complete and unreported data would lead to a different conclusion than that arrived at previously.

It has become fashionable for many periodicals to publish questionnaires for readers to complete and return. The responses are then complied and reported in the publication. Although these surveys may make entertaining reading, it is not clear to whom the results apply. How are readers of particular publications different from the general population? One would certainly expect very different responses on certain topics from readers of *Playboy* and readers of the *B'nai Brith Messenger.* It is not even reasonable to generalize such results to all readers of the magazine; the people who elected to respond may differ from those who did not. Many organizations report results of surveys of their customers or clients. Such surveys may be quite useful, but they indicate nothing about individuals or organizations that are not customers or clients.

The question of sampling and sample design—how people are selected for participation in a survey—is a critical issue for the evaluation of data because it deals with the question of generalizability of results. It is also important to determine who responded and the response rate. **[p. 26 ↓]** A survey with a response rate of 80% is certainly more credible than one with a 5% response rate. Given that a result was obtained from a particular study, can that result be considered representative of some larger population? What is the nature of that population? All too frequently one finds that it is impossible to identify that larger population.

A description of the sampling procedure is always necessary when evaluating the usefulness of data. For example, suppose that it was reported by an independent research firm that 60% of subjects given a test drive of both a Honda Civic and a Ford Escort chose the latter on the basis of overall quality. This result is impressive for Ford but becomes questionable if it is determined that all of those sampled lived in Detroit and, furthermore, that all of the individuals worked for Ford. The sampling issue applies

not only to people but also to other units, such as time, organizations, locations, and situations. A more detailed description of survey and sampling procedures may be found in two companion volumes of this series (Fowler, 1988; Henry, 1991).

A chronic problem with much research in the social sciences is that of missing data. Data may be missing for a variety of reasons, but the most frequent is nonresponse. When obtaining information from people, it is impossible to obtain data from everyone of interest. Individuals may not be found, or they may simply refuse to cooperate. Even the Census Bureau, which is charged with collecting information about the whole population, fails to obtain 100% response rates. Obviously a 95% response rate is good, and a 5% response rate is poor, but there are no clear guidelines for discounting information as a result of a low response rate. The issue of response rate applies to both the level of the observation and the variables within an observation. For example, a survey might produce a very high response rate, but a particular item in the survey may have been left unanswered by 60% of the respondents. Such an item would need to be interpreted with caution if it were used at all.

It is often helpful to know the reasons for nonresponse when evaluating information. It is also useful to compare respondents with nonrespondents on whatever information may be available for such purposes. Some information about demographic characteristics is generally obtainable, and comparisons of respondents and nonrespondents should be reported for these characteristics.

Sampling and response rates are not the only details of the data collection procedure that should be available. Copies of measurement instruments, questionnaires, coding forms, and the like help to identify what information actually was obtained, how it was obtained, and the validity of the inferences made from the data.

[p. 27 ↓]

A common means for summarizing information about trends is the use of percentages. A large proportion of government data, as well as data from other sources, is presented in percentage form. Though this may be useful in some cases, it can also be misleading. Percentages are relative. A 10% change is quite different when the base is 100 than when it is 1,000,000. Large percentage changes often arise when

computations are based on small numbers. Managers often do not understand why last year's 400% increase in sales has dropped to a 50% increase in the current year. The reason may be simply one of an increasing base on which the percentage figure is calculated. Thus percentages are seldom particularly useful unless one has knowledge of their base.

A frequent method for summarizing differences among groups involves a transformation of percentages. This transformation produces an index number. Index numbers may be calculated in many ways, but all involve a comparison of two percentages. For example, assume that 20% of the population as a whole owns a personal computer. Among engineers, 80% own a personal computer, but only 20% of architects own one. Indices representing the likelihood of ownership of a personal computer by occupation may be constructed by dividing the percentage of ownership for each group by the percentage of ownership for the population as a whole, as follows:

$$\text{Index for engineers} = 80\%/20\% \times 100 = 400$$
$$\text{Index for architects} = 20\%/20\% \times 100 = 100$$

These numbers would be interpreted to mean that the engineers as a group are four times as likely as the general population to own personal computers, whereas architects are just as likely as the general population to own personal computers. Such indices are very useful when one is trying to present information about many groups, but note that the index is the ratio of two relative measures. Thus very high (or very low) indices may reflect small or large bases for computation. In addition, each percentage used in the computation is itself an estimate. Consequently, the error present in an index is a combination of the errors present in the two percentages used in the computation.

Another example that illustrates how important it is to reflect upon the base of an index can be shown in the construction of a price index. For simplicity, consider the following consumer price index based upon two products: Brand W water and Brand Y yogurt. Assume that the prices of Brand W and Brand Y are recorded as follows:

[p. 28 ↓]

	1991 Price	1992 Price
Brand W	$1.50/bottle	$3.00/bottle
Brand Y	$3.00/6-pack	$1.50/6-pack

These data show that Brand W in 1992 was twice as expensive as it was in 1991, whereas Brand Y was half as expensive. Assume now that two different researchers have been asked to take these data and to construct a price index. Researcher A decides to use 1991 as the base and comes up with the following calculations:

	1991	1992
Brand W (relative to base)	100	200
Brand Y (relative to base)	100	50
Total	200	250
Combined index	100	125

Researcher A, using the combined index above (which is the arithmetic average of levels for the two individual brands) shows that, on average, prices rose 25%. Meanwhile, Researcher B decides to use 1992 as the base and arrives at the following calculations.

	1991	1992
Brand W (relative to base)	50	100
Brand Y (relative to base)	200	100
Total	250	200
Combined index	125	100

From the results above, Researcher B concludes that prices, on average, decreased by 25%. It is interesting to note that both researchers worked with identical data, yet arrived at dramatically different conclusions. By simply changing the base year, prices were made to appear to go up as opposed to going down.

Now Researcher C comes along and proclaims that both of the other two researchers are incorrect. He states that it is obvious that if one were to buy a bottle of Brand W and a six-pack of Brand Y in either year, the cost ($4.50) would be identical. That is, there has been no change in the total at all. Now an individual who relies upon a logical conclusion from these data can proclaim that prices have risen, declined, or remained the same, depending upon the choice of researcher! Of course, the fallacy **[p. 29 ↓]** in all of this is that the percentage relatives within Brands W and Y are not percentages based upon the same quantity. A 100% increase in the price of Brand W would represent $1.50, whereas the same percentage increase in Brand Y prices would

represent $3.00. For an effective index to be constructed, the respective percentage changes must first be weighted in proportion to their relative base prices. The weighted percentages thus are expressed to a common base (e.g., 1991) and can now be combined mathematically for averaging purposes as follows:

	1991	1992	% Relative	Weight	Product
Brand W	$1.50	$3.00	200	1	200
Brand Y	$3.00	$1.50	50	2	100
			Totals	3	300

The average product in the example above is 100, which represents the combined index number. Hence support for the conclusion of Researcher C is found, as prices have indeed remained unchanged. Sometimes when data are transformed, the transformations themselves are transformed. In fact, an index is really a transformation of a transformation—first raw data are changed to percentages, and then a ratio of percentages is obtained. It is important to understand what specific transformations have been performed and how these transformations were done. The lesson to be learned from all of this is that one must always question the ways in which conclusions are derived in the interpretation and usage of secondary information.

Similarly, any experimental or field procedure should also be described in detail. For example, in a study of consumer reactions to a new product, it would be useful to know whether the product actually was used by the consumers or whether it simply was described to them. Reports on the technical performance of products should specify the conditions under which measures were obtained. The automobile mileage estimates disseminated by the Environmental Protection Agency are obtained under conditions quite different from those under which most automobiles operate.

When evaluating the procedures employed in collecting information, the crucial question is one of bias. Was something done (or not done) in the study that would lead to a particular result, produce results that may not be generalizable, or confound the interpretation of results? Such information is not always available. When it is available, a more useful assessment of the data provided can be done. When it is not, a healthy skepticism is in order.

[p. 30 ↓]

Finally, evaluation of how data were obtained should include attention to the quality of the data themselves. Is there evidence that the measures were reliable, valid, and complete? Some measures may be very ad hoc, whereas others may be based on careful development and application over long periods of time. Some measures may reflect only a portion of the data of interest. For example, electronic scanner data may reflect the purchases of a household in a supermarket, but miss any purchases of similar products that might be purchased in a discount store or other outlet.

HOW CONSISTENT IS THE INFORMATION WITH OTHER INFORMATION?

When data are presented by multiple independent sources, one's confidence in those data is increased. Given all of the problems that may be present in secondary data and the frequent difficulty with identifying how the data were obtained, the best strategy is to find multiple sources of information. Ideally, two or more independent sources should arrive at the same or similar conclusions. When disagreement among sources does exist, it is helpful to try to identify reasons for such differences and to determine which source is more credible. This is not always easy, even with relatively complete information. When radically different results are reported and little basis for evaluating the information collection procedure is found, it is appropriate to be skeptical of all of the data.

A NOTE ON THE INTERPRETATION OF NUMBERS

Secondary data often come in the form of numbers. Numerical data have the appearance of being “hard” data, tangible and concrete, when compared to information presented in words. Yet a number is the ultimate abstraction, with no inherent meaning. Numbers are simply vehicles for carrying information. The user of secondary data

should be comfortable with numerical data, but should also understand that numbers are no better than the information they represent and the process by which that information was generated. Unfortunately, many secondary sources do not provide the most useful numerical information, as the examples above in terms of indexes revealed.

[p. 31 ↓]

Other descriptive statistics can also pose problems. Means are seldom useful without accompanying information. Generally, one would also like to have an indication of the variability of the sample or population and the number of observations on which the mean was computed. Such information facilitates the identification of significant differences, gives a better perspective on the underlying form of the data, and (other things being equal) improves the confidence one places in the data.

As an example, consider the table below, which shows the profit and losses (as a percentage of sales) of a company's six divisions for the years 1991 and 1992:

Division	1991	1992
A	+28%	−2%
B	−15%	+60%
C	+2%	−5%
D	−12%	+20%
E	+26%	−14%
F	+19%	−11%
Totals	+48%	+48%
Mean	+8%	+8%

Reporting from the above data that the average profitability across divisions for the company is the same for the 2 years is problematic for two reasons. First, the percentage increases (or decreases) reported by each division cloak the fact the some divisions had greater sales than others, and hence these divisions should be weighted more strongly in the mean calculation. Second, although the means are the same, there has been a dramatic shift between divisions in terms of profitability. That is, each division that was profitable in 1991 was not profitable in 1992, and vice versa. This example again illustrates the difficulty in relying on reported secondary summary statistics without knowledge of the data used in their construction.

SUMMARY

All data are not created equal. When using secondary sources, it is important to evaluate very carefully the information presented, to weigh potential biases, and to adopt an attitude of healthy skepticism. Conclusions should not be accepted at face value simply because they are in print, or because the claim is made that they are based on empirical research. Evidence in support of conclusions must be evaluated and **[p. 32 ↓]** weighed carefully to determine whether such conclusions are justified. Alternative explanations for research findings should be identified and considered. Factors other than those identified in the study may have produced a particular result. Only careful consideration of the methods employed to collect and analyze the data will reveal such alternative explanations. Confidence in the conclusions of one study is bolstered when these conclusions also are supported in other studies. The use of multiple sources of information is, ultimately, the best defense against being misled.

EXERCISES

http://dx.doi.org/10.4135/9781412985802.n2