

# REFLECTIONS ON DATA SOURCES IN OPINION RESEARCH

BY JOHN W. RILEY, JR.

Taking his cue from recent developments in the technology of data collection and data processing, John W. Riley, Jr., suggests that the field of public opinion research might well pay more attention to available data for the twin purposes of raising better problems for new surveys and interpreting survey results within broader frameworks. This is his presidential address to the Seventeenth Annual Conference of the American Association for Public Opinion Research, the proceedings of which appear elsewhere in this issue.

John W. Riley, Jr., is well known to the readers of the *Quarterly* for his contributions to the sociology of communications and his various articles on psychological warfare problems. Recently, after many years as Professor of Sociology at Rutgers University, he joined the Equitable Life Assurance Society of the United States as a Second Vice President to develop a program of basic social research—a plan which has been hailed as an innovation in the sociological sciences. Some of the preliminary results of one of the studies in this program are reported in this paper.

I SUPPOSE that AAPOR presidents in preparing a paper for this ceremonial occasion typically reflect upon some characteristic contribution or emphasis added to public opinion research from their own experience. In examining my own past with some such aim in view, I came upon two types of study which might conceivably have qualified. There were a number of surveys which had been conducted under unusual military conditions, and then there were some studies which used a special sociological approach. On further inspection, however, they all appeared to have been fairly typical of the traditional pattern for our field.

But then I came upon a few studies which in retrospect gave me some pause. For example, there was an early study of the changes in leisure-time behavior in a New England community that systematically tabulated all the personal items in the weekly newspaper over a seventy-five-year period.<sup>1</sup> There was an analysis of the changing pattern of destitution in an urban area, obtained by coding and tabulating a sample of the "One Hundred Neediest Cases," as reported annually in the *New York Times*.<sup>2</sup> There was also a descriptive, unsystematic use of Korean diaries and personal documents written during the 1950

<sup>1</sup> John W. Riley, Jr., "The Country Weekly as a Sociological Source," *American Sociological Review*, Vol. 3, 1938, pp. 39-46.

<sup>2</sup> John W. Riley, Jr., "The Changing Pattern of Destitution in an Urban Area," *Social Forces*, Vol. 20, 1941, pp. 64-69.

Communist occupation of the capital city of Seoul.<sup>3</sup> And very recently there was an analysis of the characteristics of successful compared with unsuccessful life insurance agents (conducted with the Bureau of Applied Social Research), based on data from some 200,000 IBM cards in the Equitable files.<sup>4</sup>

All these—and a number of others—had one element in common: the data were already available in usable form; as a researcher, I was spared the initial step of going out to collect new data.

In ruminating upon such examples, I asked myself: Does this use of available data constitute a unique research method—to be distinguished from the use of new data? Procedures for collecting new data by questioning (or by direct observation) have been clearly formulated. Much distinguished work by AAPOR members has defined the special nature of questionnaire data. In a variety of ways they have shown how these data reflect the research instruments and questioning techniques of the particular study, and specified ways in which the interpretation must be adjusted to fit the peculiar character of these questionnaire data.<sup>5</sup> But are there any comparable definitions and rules for the use of available data?

Of course, it was at once apparent that there are specialized rules for handling *subtypes* of available data—the rules of demography<sup>6</sup> or historical analysis,<sup>7</sup> for example, or (especially noteworthy in opinion research) the clearly articulated and widely used body of rules called content analysis, as codified by Berelson. But my own examples did not quite fit into any of these, even into content analysis. (One of them did not meet the quantitative criterion of most content analyses; the concern of two others was not with the communications process as such, as is typically the case in content analysis,<sup>8</sup> but with the social behavior and social structure which channel and shape opinions and attitudes.

<sup>3</sup> John W. Riley, Jr., and Wilbur Schram, *The Reds Take a City*, New Brunswick, N.J., Rutgers University Press, 1951.

<sup>4</sup> See the unpublished report by Coy G. Eklund, Agency Vice President of the Equitable.

<sup>5</sup> See, for example, Herbert H. Hyman *et al.*, *Interviewing in Social Research*, Chicago, University of Chicago Press, 1954; Samuel A. Stouffer, *Measurement and Prediction*, Princeton, N.J., Princeton University Press, 1950; or P. F. Lazarsfeld and M. Rosenberg, editors, *The Language of Social Research*, Glencoe, Ill., Free Press, 1955.

<sup>6</sup> See Philip M. Hauser and Otis Dudley Duncan, editors, *The Study of Population*, Chicago, University of Chicago Press, 1959, Chap. 3.

<sup>7</sup> See the methodology of Sigmund Diamond, *The Reputation of American Business Men*, Cambridge, Mass., Harvard University Press, 1955.

<sup>8</sup> See Bernard Berelson, "Content Analysis," in Gardner Lindzey, editor, *Handbook of Social Psychology*, Cambridge, Mass., Addison-Wesley, Vol. I, 1954, pp. 488-489; also Bernard Berelson, *Content Analysis in Communication Research*, Glencoe, Ill., Free Press, 1952.

And one of them did not start with raw data at all but with materials already on punch cards.)

Underlying such specialized sets of rules, are there, then, any general rules applying to available data? Can we state where, why, and how to tap existing materials, rather than to develop new ones? Is this a method of public opinion research with its own distinctive procedures for analysis and interpretation, and its own peculiar advantages and limitations? If so, are we paying sufficient formal attention to its potentialities in our field today?

## I

In order to begin to answer such questions, I want to explore with you some of the possible differences between available data and new questionnaire data obtained by the researchers themselves. Let me start by telling you something about a recent study that uses both types of data. (This is an unpublished study in the Equitable basic social research program, made by the Rutgers Research Group, under the direction of Matilda White Riley.) The problem under investigation was how, in a changing society, the younger generation learns certain attitudes, specifically, how they develop opinions about whether or not the wife should work. The Rutgers research may be considered in three phases.

Phase I pulls together and re-examines the relevant *available data*. Materials (originally compiled by the Bureau of the Census and the Bureau of Labor Statistics) show for the country as a whole that (1) the rate of labor force participation among married women has almost doubled in the last two decades (from 17 per cent in 1940 to 32 per cent in 1960),<sup>9</sup> and (2) among married women, the participation rate (in a given year) increases with education, so that 26 per cent of the wives with elementary school education work, in contrast to 39 per cent of those with four years of college or more. This knowledge of the broad societal trends enabled the researchers to focus the questions to be asked of the new data.

In Phase II, accordingly, new questionnaire data were collected from a sample of adolescent girls and examined in relation to this apparently critical variable—education. And, sure enough, the new data did show that those girls planning higher education were more apt to look with favor upon the idea of a future job (combined with marriage) for themselves. That is, the new generation is being socialized to attitudes which tend to reinforce the societal trend. In addition, the new study obtained data to explain this relationship, showing that higher education is associated with the attitude that the woman's job is inherently

<sup>9</sup> *Statistical Abstract of the United States, 1961*, Bureau of the Census, p. 213.

desirable, while lower education goes with a primary emphasis on the income that can be produced from a job.

But this was not all that the new study showed. An unanticipated finding also developed: The data revealed no difference in adolescent attitudes according to family economic status! That is, girls with high-status fathers and girls with low-status fathers were equally likely to want to work themselves. Acceptance is independent of family status. This puzzled the Rutgers researchers. If acceptance increased with education, why not also with economic status?

On further investigation, they were able to disentangle two quite different tendencies from the questionnaire data. High status and low status produced similar acceptance but by different paths: Higher-status fathers are more likely to educate their daughters, and this, as we have just seen, is associated with a given level of acceptance of the norm of women working. For lower-status families something else is happening that is not related to education at all. In these families, as one would expect, the mother is more apt to be working. The experience of having a mother who works seems to yield approximately the same level of acceptance of the idea among the girls.

Now such a finding, if it could be taken at face value, seems highly provocative. The Rutgers group feels it might mean that, in the process of socialization, the entire definition of the woman's occupational role is changing, that the role is no longer economically determined. Working women may no longer be tied to the stereotype of the low-income family and the inadequate male provider. The group even thinks that, as a consequence of an emerging expressive definition, the role may gradually fit more appropriately into the woman's *primary* social-emotional role within the family. But any such interpretations seemed much too grandiose to weave from the slender materials of this small local study.

And so the researchers went on to Phase III. That is, they went back to the available data for the country as a whole, to see whether these might verify and explain the survey finding. This time, of course, they looked for the economic status factor and found two scattered sets of data that seem to parallel the adolescent findings. First, they found a table that cross-tabulates two independent variables—the wife's education and the husband's income.<sup>10</sup> This table is a beautiful example of two factors clearly working at cross-purposes. It shows that, at each income level, the wife's employment rate increases with her education (this is the trend already noted in Phase I). But it also shows that, at each level of education, the wife's employment rate decreases with her

<sup>10</sup> "Family Characteristics of Working Wives," *Current Population Reports*, P-50, No. 81, Bureau of the Census, March 1957.

husband's income. Education favors employment, but high income inhibits it. What happens, of course, is that the better-educated women marry the financially more successful husbands. And the result seems to be that, at the present stage of transition, many wives are under marked cross-pressures.

This table led the researchers to expect that, gradually, over time, one of these factors will take precedence over the other. And the new questionnaire data on socialization suggest that education is becoming more powerful than income. Accordingly, they looked for further available evidence on the character of the change and found some trend data which show that, over the last decade, the importance of the income factor has indeed been declining—the differences in wife's employment due to husband's income have been growing smaller.<sup>11</sup> (Incidentally, this particular example is interesting because it challenges a common assumption of many opinion surveys, that education and income are similarly related to most dependent variables and may therefore be combined, without specific testing, in a composite index of socio-economic status. But, in this instance, education and income are pulling in opposite directions within the same family.)

This example illustrates schematically the procedure of working back and forth between available data and new data. (To be sure, the actual steps in such research never have quite the elegance of such *ex post facto* formulations of them.) Let us now consider what the available data add to the study. What are their peculiar advantages, in this instance, in comparison with the new data?

In particular, it seems to me, these census materials contribute in a number of respects which, taken together, serve to extend the scope of this analysis:

1. The census data include *additional variables* which are relevant to the analysis (economic status, in this example, including both occupation and income of the husband). That is, the available materials round out the picture, give it greater depth.
2. They increase the degree of *refinement* with which each variable is measured. (Education is calibrated to the several years of schooling in the massive census analysis but is reduced to a trichotomy in this smaller questionnaire survey.)
3. The available data greatly extend the findings of the survey *over space*. They show the extent to which questionnaire findings from a single area may be consistent with information for the country as a whole; they help to establish the generality of the survey findings.

<sup>11</sup> Jacob Schiffman, "Family Characteristics of Workers, 1959," *Monthly Labor Review*, Vol. 83, 1960, pp. 828-836; and "Marital and Family Characteristics of Workers," *Monthly Labor Review*, Vol. 84, 1961, pp. 361, 363.

4. Perhaps most interesting, they extend the findings back *over time*. That is, if the new study deals with a static cross-section, as if time were asked to stand still for a moment while we take a quick look at the state of public opinion, the census trends locate this cross-section within the larger pattern of process and change. It is because of the trend data that the dynamic aspects of the transition become apparent in the analysis.

Thus, in this particular example at least, the existing data do not merely supplement the survey data by increasing their quantity: the existing data are, rather, different in kind; they serve an additional function in the research process by extending its scope in various ways, enhancing the researcher's perspectives and his understanding of the problem at hand. In a moment, I shall want to test out this rather exciting idea by going beyond this one little example.

But, first, let us note the analytic interplay here between existing data and new data, with its implications for the ultimate scientific objective of bringing research operations and theory building into alignment. The initial use of available data in the study, you will remember, was to develop a preliminary conceptual model of what might be happening; that is, the available data were used in advance to formulate the theoretical problems for the new research. Then again, after the new data had been collected and analyzed, further recourse was made to existing materials, this time in order to interpret fully and develop the larger theoretical implications of the new research findings. So it appears that the census data were (at least potentially) useful here not only in extending the scope of the analysis but in stimulating the advance formulation and the later testing and revision of a conceptual model.

## II

Are such advantages generally inherent in the use of existing, rather than new, materials, or are they specific to the particular example we have been considering? In this example, of course, the new data were from a rather restricted sample, whereas the census data were reliably compiled on a broad base and over time. In general, however, it is obvious that the data which happen to be available on given topics will range enormously in scope and in theoretical potential.

Nevertheless, one advantage appears to belong exclusively to data from existing sources—the possibility of extending the analysis back into the past. Past events can no longer be observed directly by the researcher, nor can they ordinarily be learned through questioning beyond the recollection of respondents living today. Thus, the impor-

tant analysis of historical events or of long-term trends depends upon the prior existence of relevant materials.

But what of the other advantages from our example—the widening of scope through space, the possibility of multiplying variables, the increase of refinement in the measurement of each variable? Each of these desiderata *might* be met through new data; many surveys accomplish all of these ends. But note the appropriations of dollars, time, and research skill required! Within a given budget, it seems clear that such advantages may be *more efficiently* attained, whenever appropriate data are already at hand, by obviating the expense of massive data collection.

Let me allude to a study of family roles and their related attitudes, which will illustrate this possible extension of scope without commensurate expense. Various studies, such as those of Bales and Strodtbeck, had shown that small groups tend to develop two types of leader—an instrumental leader, who gets the group task done, and an expressive leader, who looks after the social-emotional problems of the group. The researcher in this case (Morris Zelditch) wanted to know whether this is also true of the family—whether there is some inherent tendency within the family so that, no matter where the family is, there is typically an instrumental leader (the husband) and an expressive leader (the wife). Obviously, such a universal hypothesis requires data about a wide variety of societies and cultures. The collection of new data would require not only very large financing but also considerable time and an array of anthropological skills. Accordingly, the research was done—relatively quickly and inexpensively—from existing materials, in the following way.

From lists of all societies for which ethnographers had reported relevant data on the family the researcher drew a sample of fifty six of these societies. He then proceeded, systematically, to reanalyze the data, working in an American library. The method enabled him to code each society according to its predominant family structure, and then to count the number of societies in which his hypothesis was supported. The report shows that the data do appear to support the hypothesis, and serves as a good illustration of the use of available materials to extend at a minimal cost the geographical scope—hence, the generality—of research findings.<sup>12</sup>

Working from such examples, then, we have come upon two advantages of using available data: (1) extension over time and (2) other extensions of scope—possible, but often too costly to be efficient, in

<sup>12</sup> Morris Zelditch, "Role Differentiation in the Nuclear Family: A Comparative Study," in Talcott Parsons and Robert F. Bales, editors, *Family, Socialization and Interaction Process*, Glencoe, Ill., Free Press, 1955.

new questionnaire surveys. In these examples, the data were originally obtained by researchers for research purposes; their advantages over new materials accrued largely because they had already been assembled—in depth, in breadth, and over time. Now let us consider other available data which differ in kind, for example, the rich materials which, though highly useful for research, were not initially produced for any research purpose at all. Consider the letters used by Thomas and Znaniecki in their classic study of Polish peasants.<sup>18</sup> You will recall that the authors in this case were studying the changes that took place in the belief systems and values of the Polish family when the traditional equilibrium was disrupted by the emigration of the father to the United States. The situation is ideal in that the letters represent the entire communication which took place between the father and the rest of the family. They provide an insight into the private, intimate life of the family and its continuous development over time that would be difficult to obtain through questioning or observation. That is, here the available data seem to tap a distinctive *type* of information.

Besides letters, one can think of diaries, documents, sermons, novels (often used in content analysis), as well as the abundant (but less fully utilized) case records in the fields of medicine, law, insurance, and social work. In addition to these verbal data, such nonverbal materials as works of art, pieces of music, scientific inventions, various artifacts, can be used as Sorokin, for example, has used them, as indicants of changing value systems over centuries of Western culture. Such widely varied sources of materials clearly open up whole new content areas for study, areas often inaccessible to questioning even in the most sophisticated forms. Thus it appears that the use of available materials may have another unique asset as a method—it may transcend some of the content limitations imposed by the questionnaire as a data-collection instrument, reaching out for new kinds of information about the topic under study.

### III

Let us consider further, apart from such advantages, whether the use of available data has distinctive characteristics as a method. We shall look briefly at the forms in which the data may be obtained and at the procedures for handling such data.

In some respects, the raw materials already at hand are methodologically quite similar to those newly obtained from questionnaires. As verbal materials, they may occur in either structured or unstructured form, and they require coding, analysis, and interpretation as in the

<sup>18</sup> William I. Thomas and Florian Znaniecki, *The Polish Peasant in Europe and America*, Chicago, University of Chicago Press, 1918.



usual survey. They may be handled either quantitatively—or descriptively, as in the analysis of the Polish peasant letters—or they may be developed into ideal types, as Max Weber did in his comparative studies of religion. In other respects, however, they are strikingly different. Works of art or pieces of music, for example, do not consist of words and hence require special handling. Or, again, the biases and limitations of family case records or of correspondence between Polish *émigrés* and their families—of data *not* produced for research purposes—are different from interviewer bias and require their own procedures for handling in analysis.

Available data, then, may have quite a distinctive character, even in their raw form. Beyond this, very unlike the usual new data, they may appear in several other forms—indeed, at any of the several stages of the research process itself. Thus, they may be coded and punched and ready to be tabulated like the “data bank” cards at the Roper Public Opinion Research Center,<sup>14</sup> or the One-in-a-Thousand Population Census Sample which will soon be available on either cards or tapes from the Bureau of the Census. Or they may be already tabulated and ready, in statistical form, for analysis or interpretation, as with published census reports. Or they may be the findings of completed studies available for use in reinterpretation or reanalysis, as in the Human Relations Area File. Already in our own field we have some impressive illustrations of different types of secondary analysis, ranging from Bruner's *Mandate from the People* and Almond's *The American People and Foreign Policy*, to such relatively recent works as Roper's *You and Your Leaders*, Campbell *et al.*, *The American Voter*, and such codification studies as Hyman's *Political Socialization* and Klapper's *The Effects of Mass Communication*.

That data may be available at all these research stages points to still another unique feature of the method: in the analysis the researcher must try to reconstruct the process by which the data were originally assembled by somebody else. In handling data which he has himself collected, he knows this process. He is aware of the limitations and possible biases of his materials; he can evaluate their validity as indicants of the concepts he has defined. Consequently, he can adapt his analysis and his interpretation to the nature of the data themselves. Such adaptation is equally important when he did not collect the data, but the means of assessing these data may be difficult, sometimes impossible, to obtain.

And this point, of course, suggests some important hazards and

<sup>14</sup> Note the recent report by Ithiel de Sola Pool and Robert Abelson in which they utilized data from over 100,000 interviews from some sixty-six surveys, “The Simulatics Project,” *Public Opinion Quarterly*, Vol. 25, 1961, pp. 167-183.

problems in the utilization of available sources. Because the data were not originally brought together for the purpose at hand, they are often incomplete or in a form in which they are not readily usable. And frequently, of course, there are no data at all which are relevant to the theoretical model under study.

I need, however, to make one final comment. Today the collection, classification, and storage of excellent materials is increasing rapidly in very many areas of interest to public opinion research. And, perhaps more importantly, the development of high-speed computers for storing, processing, and retrieving data is already contributing to the feasibility of handling and analysis. Thus, the time seems ripe for more creative use of these materials, both to raise better problems for new surveys and to interpret survey results within a broader framework.

This calls to mind the account of one of the earliest social surveys to combine available data with questionnaire data. When Charles Booth, late in the nineteenth century, wanted to document the conditions of life on London's East Side, he all but discarded the official statistics as being too aggregative. He needed, according to his accounts, supplementary materials, and he proceeded to gather new data himself. Today, the shoe seems to be on the other foot. We may have come full circle since 1886, when Booth started to compile his data for the *Life and Labour of the People of London*. Perhaps one of our needs in 1962 is to place *more* rather than less emphasis upon data already at hand.