THE QUALITY OF SURVEY DATA

Barbara A. Bailar, U.S. Bureau of the Census

I. Introduction

Quality is an "in" thing these days. Manufacturing companies are putting quality assurance systems in place that emphasize "doing it right the first time." Claims are made that an increase in quality goes hand in hand with an increase in productivity and a decrease in cost. This seems counter-intuitive to many, but the manufacturing companies that are following the ideas of Deming, Juran, and Crosby say that this is true. Can we apply some of these same kinds of techniques to improve the quality of survey data?

This leads to an interesting question. What do we mean by quality? How do we measure it? What means do we use to say that one survey or census is better than another? Quality is difficult to define, but what is immediately apparent is that quality is multi-dimensional, is perceived in different ways, and is intimately linked to the way the end product is to be used. The quality of survey data is also built in (or destroyed) at various interlocking steps in the production of the data. Just as one can visualize the assembly-line production of manufactured goods, so can one visualize the assembly-line production of survey data. We might think of it in terms of the discrete stages of design, collection, processing, and interpretation. Poor quality at any of these stages is carried through to the final data product.

The concept of statistical control, as discussed by Shewhart and and others implies that a process is reproducible, that the measurements resulting from a process are the product of an identifiable statistical universe, and that there is standardization of the process when carried out by different people and in different places. To be able to detect whether a process is in statistical control, one must be able to identify certain quality characteristics of that process.

II. Quality Characteristics

One way to arrive at a set of quality characteristics is to imagine what we would like to have happen in an ideal situation. If one didn't need to worry about cost and time restraints, what are desirable survey characteristics? All things considered, one would opt for the following:

- Frame that covers the complete universe with no ambiguity.
- Probability sampling.
- Conceptual clarity.
- Operational definitions that fit concepts.
- Sample design perfectly executed.
- Most knowledgeable respondents always report.
- Respondents report accurately and completely.
- Data are recorded accurately.
- Data coded accurately.
- Data weighted accurately.
- Data reported not replaced in processing.
- Sampling variances computed accurately.
- Sampling variances small.
- Verification of interviews shows little or no inconsistency.
- Tabulated results show consistency with other data.
- Inferences based on data are supported by measures of uncertainty in data.

There never has been a survey, nor will there ever be one for which there are no compromises. Rarely do we ever have a universe listing for which there are no ambiguities; however, in the best situations we are confident that the elements not included in the universe listing are small in number or in size or both.

Most of the surveys on which public policy is based are probability samples, but that is not true for some of the industrial surveys. In these surveys, a cut-off sample is often used in which large industrial establishments over a certain size are included with certainty and smaller establishments are omitted completely.

In the design stages, careful work is usually done on trying to clarify concepts. The unemployment concept is one which has received much attention and gets reviewed periodically by Presidential Commissions. Not only is the concept reviewed but its translation into operational definitions and questions is reviewed. However, for surveys in general this is probably an area in which much more work needs to be done, because the translation of concepts into definitions and then into questions is very difficult. Simple pretests often illustrate weaknesses.

Sample designs are usually not perfectly executed because of a variety of circumstances related to the fallibility of interviewers, the Post Office, or the telephone companies; to the inadequacy of maps; and to other factors.

The most knowledgeable person for a particular sample person or establishment is frequently not the respondent. To save money, one respondent often reports for an entire household. In establishments, the questionnaire may be filled by people who are new to the company, may not really understand what is being asked for, do not read the detailed instructions, and are reluctant to look up records.

We tend to think of the questionnaire as presenting the same stimulus to the respondents, whether they are presented with the questionnaire through an intermediary known as an interviewer who comes in person or by telephone or whether the questionnaire comes in the mail. We know that interviewers are not just passive survey instruments. They often determine whether or not a unit responds; they interpret the questionnaire; they represent the survey organization; they record the data. In cases where there are no interviewers, the format and appearance of the questionnaire take on an added burden of representing the survey organization and the accompanying instructions for filling the form supply the interpretation. However, it is a rare survey in which the respondents report fully and accurately. Often they do not understand questions, are unwilling to make the effort to find records to provide answers, forget the answers, or are unwilling to provide them. Interviewers and respondents make errors in recording data.

Data often require coding which is carried out by clerical staff working with instructions. Items such as occupation, industry, standard industrial classification, product line, and so forth require detailed coding. Often the written responses are ambiguous or vague. Instructions are tedious and sometimes vague. The job itself is monotonous. As a result, errors occur.

Survey data then usually are subjected to various processing steps which are supposed to "clean" it, weight it, and get it in shape for tabulation and analysis. Weighting seems easy, since we tend to think of survey weights as the inverses of selection probabilities. But weighting now is used to take care of nonresponse, and coverage deficiencies. Different weights are given to individuals depending on whether they are being tabulated as persons or family members. Sometimes weights are reduced in order to dull the effects of outliers.

Again, in an ideal survey, the sampling variances would be calculated correctly, observing the sample design, and reflect the effect of interviewer and coder variability, and the effect of imputing for nonresponse. In some cases sampling variances do not reflect a complex sample design. Rarely do they include interviewer, coder, or other processing variances. In cases where cut-off sampling is used, sampling variances are not computed because there is no probability sampling.

One hopes for small sampling variances, or at least no bigger than what was expected. However, if the sample design was implemented in a troublesome way, large variances can occur.

As data are verified in some kind of reinterview or monitoring program, one looks for consistency of response. However, inconsistency does occur if a question is ambiguous, if proxy respondents are used, or the respondent changes responses for no good reason.

As the tabulations become available, consistency with other data is of concern. Often there are past reports from the same establishment, or another body of data on the topic are available.

Finally, in the interpretation of the data in the survey report, it is important that the inferences made are supported by the data. If differences between groups are discussed, then the sampling error of the difference must support the idea of a real difference. However, just looking at sampling errors is not enough. One should also take into account nonresponse, measurement errors, and degree of editing.

So, in practical terms, these ideals are hardly, if ever, realized. How then do we decide on the quality of a survey? The following list seems to contain what we use as proxy indicators for data quality:

Coverage ratios of sample totals to estimated universe totals.
Response rates for sample units.
Response rates for questionnaire items - both weighted and unweighted.
Edit failure rates.
Consistency with past reports.
Amount of reported data replaced.
Confidentiality checks.

Many of these quantities are available, if at all, only to people within the survey organization collecting and processing the data. Of the items listed above, sampling variances are the most frequently reported in publications. With this paucity of information, it is difficult to understand how users of survey data measure quality.

However, since some of these data are available to survey organizations, perhaps they make use of it to measure quality or to improve quality. Often, the data are used as indicators of the presence or absence of problems. For example, a high response rate is viewed as a measure of high quality and the absence of a problem. However, it is also true that these proxy indicators focus mostly on the collection phase of survey problems. Table 1 shows the proxy indicators by the four stages of surveys.

Though many of the indicators apply to design problems, they are more often used as indicators of problems in the collection phase. For example, a high item nonresponse rate is frequently looked on as an interviewer problem, but it may be more a problem with the question itself. In fact, a high edit failure rate for a given item along with a high nonresponse rate for the same item may be a real signal that the question needs work. Also, although three of the items can be used as indicators of the quality of processing as well as collection and design, they are more often used as indicators of quality of collection.

Notice that none of these quality indicators applies to the interpretation of data. This signifies that a separate step is necessary in the review of a report to measure the quality of interpretation.

Most of the indicators, if they exist, are used singly rather than in combination. If they were used together, they would be much more likely to spotlight the real problem, since these indicators are rarely due to only one cause.

For example, if we were to construct cause-and-effect diagrams as espoused by Iskikawa using the various indicators as the quality effects, one would see overlaps on the causes contributing to quality characteristics. Table 2 is an attempt to show the various causes for the indicators of poor quality.

When people compare unit response rates as measures of quality of a survey, they are comparing the ability of interviewers or analysts to get respondents to answer questions which may be presented by mail, telephone, or in person. We often see comparisons of the response rates of random-digit dialing telephone surveys with those of personal interview surveys. The response rates of RDD surveys are lower. We usually attribute this to a difference between collecting data by telephone or in person, but it really is a difference in the entire stimulus to a respondent. The interviewer-respondent interaction is very different. Questions may seem more complex on the telephone. We know that different populations are accessed. Therefore to understand why the response rates are different, much more needs to be understood about the interviewers, the respondents, and the questionnaire.

There are several paradoxes in the use of quality indicators. Some illustrations show why they need to be used together instead of singly. The first example will focus on response rates and coverage rates in the Current Population Survey (CPS). The CPS is recognized as a "high-quality" survey which has received much attention over its 40-plus years. The response rates are generally high, in the neighborhood of 95 to 96 percent. Monthly data are produced regularly on response rates and on coverage ratios. By coverage ratios, I'm referring to the ratio of weighted population totals for various groups to independent estimates of population totals for those same groups. What can we learn from an examination of response rates and coverage ratios?

Let's look at response rates first. We know the average nonresponse rate is low. Is it uniformly low over all months? Is it uniformly low over all states and substate areas? Is it uniformly low by month in sample of the respondents? Could control charts help us in spotting problems? To answer these questions, I took data on nonresponse rates for a 15-month period from October 1981 to December 1982.

Figure 1 shows plots of the noninterview rates which include refusals, not-at-homes, and persons temporarily absent from their residences, as well as a residual category of "other" for this time period. The standard errors of these monthly rates were calculated using observations from the 50 states, the District of Columbia, and several substate areas such as New York City, Chicago, balance of Illinois, and so forth. Figure 1 also shows plots of the refusal rates for the same time period. The rates are shown in Figure 1 with ± 3 standard errors added to the rate.

The first thing that shows is that the March noninterview rate and refusal rate is different from the others. It is higher and the interval is wider, showing that there is more variability in the rates in March. This is not a new finding. The March interview is very different from the other months, because it includes a very long population supplement in which data are collected on fertility, migration, and income. The CPS interviewers know very well that labor force data are the most important data to be collected and they do not want to jeopardize the data.

So the answer to the question about whether the nonresponse rate is uniformly low over all months is yes (even in March the nonresponse is less than 5 percent) but that March represents an unusual point. Notice that March is higher than July or August, and that the higher nonresponse rate also shows in a higher refusal rate.

To answer the question about whether the nonresponse rate is uniformly low overall states and substate areas, Figure 2 shows for March 1982 the response rates for states and selected substate areas. The plot shows that there is substantial variation, with over half of the points falling outside the 3 standard error range. For the most part, one sees the same areas as outliers month after month.

Some interesting observations can be made from the chart. The data points for substate areas are plotted on the same line as for the state. Thus the first set of seven points on the leftmost line refer to California and its sub-state areas. Notice the range of variability in March, San Francisco at 7.43 and San Diego at 2.78.

One is tempted to say that there are really two different populations represented here. One might consist of more urban areas, and thus we see New York City, Chicago, Dallas, San Francisco, and some others very high. However, this pattern does not hold. Many times the balance of Colorado has a higher nonresponse than Denver. Baltimore generally is not beyond the upper bound of the chart, while the suburbs of Washington, D.C., that lie in Maryland are higher than anything else in Maryland. Michigan frequently has a higher nonresponse for the remainder of the Detroit SMSA than it does for the City of Detroit, and the balance of the state of Michigan is generally high. Some states are also high, such as Arizona, Hawaii, and Alaska.

Thus, the answer to the second question is that there is not uniformity among the response rates for states and substate areas. This leads to a question about the uses of CPS data. If nonresponse is an indicator of bias, then states and substates with higher nonresponse rates are probably subject to more bias. If these areas are competing with each other for funds based on data from the CPS, are they treated equitably when biases are not held constant? Should we insist on equal response rates for competing areas?

Another question that arises is whether or not the type of area accounts for the nonresponse rates or do the interviewers. Probably both have a part in this, but evidence from this study as well as several others shows that interviewers are probably the biggest factor.

Figure 3 illustrates the nonresponse rates for the city of Chicago. Frequently the rates for Chicago were higher than for any other place for which the rates are tabulated. They are always far above the national ratio plus 3 standard errors. Even this wouldn't be so bad if the pattern were steadier, but the rate of 13.70 for August 1982 was way beyond what one could expect. Were interviewer assignments changed in May and August?

Several other questions could be asked, but the use of control charts by area can be useful. There are data available to develop control charts for each area and then they can be generated each month. Data are also available for individual interviewers. In the spirit of Deming, it might be useful to distribute such charts to the interviewers and their supervisors so that each person's impact on the survey can be seen and controlled.

These few examples barely scratch the surface of what we could learn from control charts set up for nonresponse rates. It is clear from these examples that we cannot predict future performance very well from past data. Control charts would help us see what is happening and understand the reasons for it.

What do we know about item nonresponse for labor force items and its effect on the data? Once a CPS interviewer is inside the door, the data are gathered on the labor force items. The interviewers know the importance of the survey and that labor force is the topic of most interest, so that very few nonresponses occur for labor force items. In a recent inquiry, I learned that from records for approximately 130,000 persons, only 50-60 persons required imputation for labor force. Obviously, item nonresponse, occurring from skipping items is not a problem in the CPS.

But what do we know about the coverage for CPS and the resulting weighting of data for the respondents to take care of the missing people?

In a way this is also a method for handling nonresponse, but the nonresponse is of a different kind. Since we believe that the sampling frame adequately covers the universe, there must be incomplete recording of the persons within households. This is a type of item nonresponse, where the item is the list of all persons in the household. Thus, we may think of some of the weighting of the CPS data as a procedure for handling nonresponse.

CPS households have a base weight reflecting the probability of selection in the sample. There is a noninterview adjustment which adjusts for the nonresponding units. There is a second-stage ratio adjustment which inflates the population to independent population controls, which reduces both bias and variance. If we look at the ratio of the weights after the second-stage adjustment to the base weights, we would get the effect of the nonresponse adjustment for coverage. These ratios are not trivial. The overall ratio for the U.S. as a whole in January, 1983 was 1.15 meaning that the final weights for the responding persons were 15 percent higher than the base weights to take care of the problem of coverage. The range for these ratios was very large, and the more rural parts of the states or rural states seemed to have very high ratios. Table 3 shows the ratios by state and substate areas. Very few of the ratios are trivial, indicating a need to adjust the data for coverage. And though the need to adjust data to compensate for item nonresponse is small, in a sense this weighting does adjust for the incomplete roster of the group of people who live in CPS sampled households.

These adjustments are very large for some states and for some groups, such as blacks in certain age groups. This is an indication that the quality of the data for certain states is lower than for most states and that the quality of the data for blacks in certain age groups is less than that for whites. This is not to say that there is not an effort to compensate for missing data, but the reweighting of data rests on the assumption that the missing people are like those who responded, within a given age, sex, race category. This may or may not be true, and to the extent that it isn't, affects the quality of the labor force statistics.

A second example to show complex interactions is in the Annual Survey of Manufactures. This survey, as its name implies, is carried out yearly to provide a range of information on the nation's industrial enterprises. The survey form is sent by mail to a sample which includes all of the very largest establishments, none of those with total employment less than 5, and a selection of those in between these extremes. The unit response rate is about 85-90 percent, with an aggressive follow-up of the largest establishments and less concern with smaller establishments which do not contribute as much to totals. Data for the smallest firms are imputed from administrative records -- tax returns --for employment and payroll and imputed on the basis of industry ratios for other items. There is some doubt that the use of industry ratios for these small firms is a good method of imputation.

The data supplied by the respondents are put through a computer edit in which blanks are filled in and inconsistencies eliminated. Table 4 shows the results of the computer edit and the analyst for some key items for one industry, motor vehicles and car bodies. It should be noted that supplying values from administrative records was not counted as an impute or a change. Note that only for salary and wages were at least 90 percent of the establishment records unchanged. However, the records that remained unchanged tended to be the large establishments so that even for an item such as cost of materials for which only 27 percent of the establishments' reports were left unchanged, 85 percent of the final tabulated value was left unchanged. The changes tend to occur in the smaller firms. This can be seen in the imputed column in which for cost of materials, 64 percent of the establishments left the item blank, but the imputed values accounted for 1 percent of the total. We also see that the computer is having far more impact on the data than the analysts. The analysts seem to concentrate their attention on the larger cases.

The example from establishment surveys points out some important quality problems. Nowhere in any publication about these surveys are the item imputation rates given. Users of the data are given the impression that all the data items are of equal quality, and all are based on the 85-90 percent of units that report. Therefore, the users are not given enough information to judge the quality of the data for the diverse uses they may make of it. Another problem is one of standards. How much imputed data are acceptable? If the Bureau is imputing 50-60 percent or more of a data item, is it worthwhile collecting and publishing the information? If the data are important, shouldn't a bigger effort go into the collection?

A final example that focuses on imputation rates is the Post Enumeration Survey carried out by the Bureau to develop measures of the net undercount of the 1980 decennial census. The Bureau used the April and August 1980 CPS samples as independent sources of persons to match into the census records to see if they were counted. Some persons were matched immediately; some addresses could not be found; some addresses were poor and locations to look in the census could not be determined. Those addresses which were not matched were sent back to the field in an effort to get more information so that a determination of whether or not the person was included in the census could be made. Thus we hoped for all the records to be sorted into two piles -- those which were matched to the census and those that were unmatched. Unfortunately we had three piles -- the two we hoped for plus a third which was not enough information to match. Those without enough information to match were imputed a match code -- either matched or unmatched, with the unmatched contributing to the estimate of the census undercount. The percent of records for a state or substate area that was imputed ranged from 2 percent for several places to a high of 15 percent for Dallas. The percent of matched cases imputed was low, on the whole, but the percent of unmatched cases imputed was very high. Central city rates tended to be higher than other places. Seventy percent of the unmatched cases in Dallas were imputed. Also, the imputation system favored imputing the status of nonmatched. It was not an even split.

The PEP has been mentioned as a tool to adjust the census data for an undercount. The uses for the adjusted data are to apportion the seats seats in the House of Representatives and to apportion money in Federal funding programs. If anyone gains anything, some one else loses. The data for each state competing against each other should be of equal quality. It is clear from the imputations that this is not the case.

From these examples, it is clear that these quality indicators must be considered together, as a system of information, not in isolation. The trade-offs that need to be made in designing surveys can be made knowledgeably only with information on the complete set of indicators. When we design surveys with equal coefficients of variation for states or other units of measurement, because we want the units to have data of the same quality, should we also insist on equal response rates and imputation rates? In addition, we need to keep an up-to-date database on interviewers and analysts who are in the position of collecting data, changing data, and writing specifications for processing data.

III. Improvements in Quality

It is interesting to note how many decisions that are made about quality are made on the basis of hunches rather than on data. For example, in the redesign of the household surveys which is underway at the Bureau, decisions were made to overlap the new CPS design with the old CPS design to reduce the number of new interviewers that would need to be hired. Arguments could be made for improvements in quality either by those favoring the maximum amount of overlap and by those favoring the minimum amount of overlap. On the overlap side was the ability to estimate variances in a much improved way. On the "maximum" overlap side was the reduction in the number of new interviewers needed and the ability to train and supervise them. There is a strong feeling that new interviewers are less likely to produce high quality data. What is the evidence on this? Do they have lower response rates, both unit and item? Do they have higher refusals? Higher edit failure rates? Are they more likely to quit and need replacement in early months? Do reinterviews show high rates of coverage and content failure? How long do these effects last? 3 months, 6 months, 1 year? Do we see differences in quality in offices with higher turnover rates than lower turnover rates? The answer in the redesign was that we knew very little. There was anecdotal evidence, but practically no hard data. No database exists which links length of experience with quality indicators. Yet almost every survey designer worries about inexperienced interviewers.

We also worry about respondents and how much they learn about a survey. At the Census Bureau, the demographic survey people worry about respondent "conditioning" and whether respondents report as accurately on later survey rounds as in early survey rounds. We have evidence that they report in different ways. They are more likely to report more people as being unemployed, as being victims of crime, or as spending more money on consumer goods the first time they are included in a survey than in return visits. The economic survey people, who generally mail questionnaires to respondents, worry about respondent "learning" and whether respondents report as accurately when they are new in a survey as when they become more experienced. There is evidence that the respondents may go through a learning process and that there may be a "shake-down" period. In one case, the early reports are thought to be more accurate. But how is this accuracy measured? On the demographic side, the axiom "more is better" carries a lot of weight so the earlier reports are thought by many to be more accurate. On the economic side, the later reports become more consistent with earlier reports so respondents are assumed to have learned how to report. There are no objective measures.

A third example about the use of hunches on quality has to do with response rates. All things being equal, one would prefer a survey that had higher response rates than one with lower response rates. But is it possible that a survey with a 90 percent response rate is better than one with a 95 percent response rate? Response rates are only indicators of potential bias but are not bias measures themselves. A 95 percent response rate can mask differential response by segments of the population, such as white and black, for which estimates are to be made and compared. If the response rate for whites is substantially higher than that for blacks, the comparisons may be misleading. A 95 percent response rate can also mask curbstoning by interviewers, substantial item nonresponse rates, high nonresponse for items towards the end of the questionnaire, high edit failure rates, and the like. By using only response rates, we tend to overlook other serious potential sources of bias.

What can we do about this? We seem to be designing surveys to minimize potential risks in the absence of data. We need to develop data that can be used to help us explicitly in the trade-offs that are always required in survey work. We need to know if an extra day's training of interviewers is worth the cost. We need to know if we allow an increase in sampling variance we can decrease the bias.

The Bureau is in the process of establishing a telephone facility in Hagerstown, Maryland, where we can experiment with the collection of data by telephone. We have several questions to be answered before we are willing to recommend that the continuing surveys conducted by the Bureau be done at a centralized telephone facility. We also have questions about whether these surveys are suitable for random-digit dialing. What we need to do is to get a better understanding of the entire system of interviewing, not just the end result. To do that we want to establish the following quality procedures.

<u>Quality circles</u>.

We want to establish quality circles in which the interviewers can discuss their problems and potential solutions. This gives us much more insight into problems that respondents have with questionnaires, in addition to providing interviewers with tips on how to improve procedures.

<u>Monitoring</u>.

We want to establish a systematic program of monitoring interviews in which a probability sample of each interviewer's work would be monitored by a person with more experience on the survey. This monitor would record events on how difficult the interview was to obtain, how the interviewer induced cooperation, probing, feedback to respondent, how often the questions were reworded, how accurately the responses were recorded, and any problems on the questionnaire. Monitoring will give us a way to target needed training on introductions, specific questions, and highlight problems with the questionnaire. Monitoring should also reveal to us why certain interviewers are more likely to get overrepresentation of certain population subsets.

Reinterview. In personal visit surveys, reinterview is used both as a tool to motivate interviewers and as a method to estimate measurement variance. In telephone surveys, monitoring can serve as the motivation factor besides offering learning opportunities, and reinterview can serve as the mechanism to estimate measurement variance. By selecting a probability sample of every interviewer's work for reinterview, we can estimate measurement variance by length of experience of interviewers, by region of the country, and by other variables. We will also be able to see which questions are more susceptible to high measurement variances.

Interpenetration of Assignments. In one sense it is easier to interpenetrate the interviewing assignments of centralized telephone interviewers than the assignment of field interviewers and in another sense it is more difficult. It is easier because it is less expensive when there are no travel costs. It is more difficult because interviewers work on shifts and weekends and probably do not access the same kinds of populations. It is highly unlikely that a telephone interviewer working a 9 a.m. to noon shift would get the same type of respondents as an interviewer working a 7 p.m. to 10 p.m. shift. Given that those problems are resolved, we intend to have interpenetrated assignment patterns. This gives us the ability to estimate the correlated component of response variance, the interviewers' contribution to the total variance of survey statistics. It lets us know immediately if some interviewers are getting much higher item nonresponse rates, different response patterns, and so forth. It makes retraining possible on the areas of poor performance.

Statistical Control Charts. Statistical control charts have been used to advantage by individual floor workers on assembly lines. They are one of the primary tools used by quality circles to suggest ideas of improving quality. Control charts should be used to plot the data by interviewer on response rates for key items, edit-failure rates for key items, reinterview results for key items, and so forth.

With the results of statistical control charts, monitoring, reinterview, and the interpenetrated assignments, we should be able to look at the quality of survey data in a much more integrated way. We should be able to use the results to identify problems in the survey question and be able to differentiate those from problems specific interviewers are having. We should be able to target retraining on specific items or techniques. We should be able to measure whether experienced interviewers collect better data, how much experience is needed, and what kinds of activities need to be carried out before that level of experience is reached.

As the survey world currently exists, the concept of quality of survey data is not clearly defined and is certainly not measured well. We tend to measure quantities that are fairly easy to collect and which we hope are related to quality. We design surveys on a combination of hunches and proxies for quality. We are starting to move to a system for which we will be able to collect more data on measures that are directly related to quality. We then need to be able to use those measures in models that explicitly allow for trade-offs in the levels of quality.

TABLE 1. -- Proxy Indicators of Quality by Stages of Survey Where Problems May Exist

	Stages of Survey			
	Design	Collection	Processing	Interpretation
Coverage ratios	x	x		
Response rates				
Unit		x		
Item	x	x		
Edit failure rates	x	x		
Consistency checks	x	x	x	
Data replacement	x	x	x	
Variance	x	x	x	
Confidentiality			x	

TABLE 2. -- Indicators of Poor Quality

Causes of Poor Quality	Low Coverage Ratios	Low Unit Resp. Rates	Low Item Resp. Rates	High Edit Failure Rates	Low Consis-tency	High Data Repl.	High Variance
Sample design inadequate	x						x
Interviewer problem							
Experience	x	x	x	x	x	x	x
Training	x	x	x	x	x	x	x
Motivation	x	x	x	x	x	x	x
Coder problem							
Experience			x	x	x	x	x
Training			x	x	x	x	x
Motivation			x	x	x	x	x
Respondent problem							
Experience	x	x	x	x	x	x	x
Motivation	x	x	x	x	x	x	x
Proxy		x	x	x	x	x	x
Questionnaire problem							
Too complex			x	x	x	x	x
Too long		x	x	x	x	x	x
Not clear			x	x	x	x	x
Skips too much			x	x	x	x	x
Instructions not clear			x	x	x	x	x
Mode of interview	x	x	x	x	x	x	x
Analyst problem							
Experience		x	x	x	x	x	x
Training		x	x	x	x	x	x
Motivation		x	x	x	x	x	x
Processing problem	x	x	x	x	x	x	x

TABLE 3. Ratios of Final CPS Weights to Base Weights-January 1983

Area	Ratio	Area	Ratio
California	1.13	Illinois	1.18
Los Angeles-Long Beach SMSA	1.14	Chicago City	1.20
San Francisco-Oakland SMSA	1.14	balance Chicago SMSA	1.11
Anaheim SMSA	1.12	balance St. Louis SMSA (Illinois part)	1.10
San Diego SMSA	1.13	balance Illinois	1.26
San Bernardino SMSA	1.11		
San Jose SMSA	1.11		
balance of California	1.12	Ohio	1.08
		Cleveland City	1.11
New York	1.19	balance Cleveland SMSA	1.11
New York City	1.19	Cincinnati SMSA (Ohio part)	1.10
Balance New York LMA	1.19	balance of Ohio	1.07
Nassau-Suffolk SMSA	1.13		
Buffalo SMSA	1.13	Michigan	1.06
balance of New York State	1.24	Detroit City	1.12
		balance Detroit SMSA	1.13
Pennsylvania	1.14	balance of Michigan	1.01
Philadelphia City	1.21		
balance of Philadelphia SMSA in Philadelphia	1.12	New Jersey	1.12
		Newark SMSA	1.15
Pittsburgh SMSA	1.12	balance Philadelphia SMSA (in New Jersey)	1.19
balance of Pennsylvania	1.13	balance of New Jersey	1.09
Texas	1.19		
Houston City	1.16	Florida	1.25
balance Houston SMSA	1.63	Miami SMSA	1.12
Dallas City	1.19	balance of Florida	1.28
balance Dallas SMSA	1.19		
balance Texas	1.12		

TABLE 3. -- Ratios of Final CPS Weights to Base Weights-January 1983 (cont.)

Area	Ratio
Massachusetts	1.21
Boston SMSA	1.25
balance of Massachusetts	1.17
Indiana	1.11
balance Cincinnati SMSA in Indiana	1.08
Indianapolis SMSA	1.23
balance of Indiana	1.08
North Carolina	1.27
Missouri	1.02
St. Louis City	1.09
balance St. Louis SMSA (Missouri)	1.07
Kansas City SMSA (Missouri)	1.08
balance of Missouri	.95
Virginia	1.11
balance of D.C. SMSA in Virginia	1.06
balance of Virginia	1.12
Georgia	1.17
Atlanta SMSA	1.22
balance of Georgia	1.14
Wisconsin	1.18
Milwaukee City	1.14
balance of Milwaukee SMSA	1.13
balance of Wisconsin	1.20
Tennessee	1.24
Maryland	1.02
Baltimore City	1.00
balance of Baltimore SMSA	1.02
balance D.C. SMSA in Maryland	1.04
balance of Maryland	1.00
Minnesota	1.06
Minneapolis-St. Paul SMSA	1.08
balance of Minnesota	1.03
Louisiana	1.15
Alabama	.99
Washington	1.17
Seattle-Everest SMSA	1.12
balance of Washington	1.22
Kentucky	1.05
Cincinnati SMSA in Kentucky	1.06
balance of Kentucky	1.05
Connecticut	1.12
Iowa	1.09
South Carolina	1.29
Oklahoma	1.14
Kansas	1.22
Kansas City SMSA in Kansas	1.25
balance of Kansas	1.22
Mississippi	1.22
Colorado	1.03
Denver SMSA	1.02
balance of Colorado	1.06
Oregon	1.23
Arkansas	1.25
Arizona	1.18
West Virginia	1.22
Nebraska	1.13
Utah	1.07
New Mexico	1.27
Maine	1.08
Rhode Island	1.14
Hawaii	1.09
District of Columbia	1.22
New Hampshire	1.22
Idaho	1.21
Montana	1.10
South Dakota	1.13
North Dakota	1.14
Delaware	1.19
Nevada	1.25
Vermont	1.15
Wyoming	1.33
Alaska	1.57
Total U.S.	1.15

TABLE 4. -- Percentage of Records Imputed and Changed by Computer and Analyst for 1982 Annual Survey of Manufactures Motor Vehicles and Car Bodies

Item of interest	% unchanged by computer or analyst		% of blanks imputed by computer		% of data rejected by computer and imputed		% of data raked to subtotal by computer		% accepted by computer and changed by analyst		% imputed from blanks and changed by analyst		% of data raked by computer and changed by analyst	
	Ests.	Total	Ests.	Total	Ests.	Total	Ests.	Total	Ests.	Total	Ests.	Total	Ests.	Total
Total employment	72	50	15	0	1	0	8	44	2	4	1	0	0	0
Production workers	24	48	64	2	0	0	8	47	1	3	1	0	0	0
Salary and wages	97	97	0	0	1	0	0	2	2	0	1	0	0	0
Worker wages	30	97	2	0	2	0	63	2	0	0	1	0	1	0
Plant hours	26	68	65	2	1	0	1	1	5	25	1	0	0	0
Cost of materials	27	85	64	1	1	0	2	5	1	3	1	0	1	1
Cost of parts	25	80	0	0	0	0	69	10	2	3	1	0	2	0
Cost of resales	27	91	1	3	0	0	64	0	2	0	1	0	1	0
Cost of electricity	25	80	0	0	2	0	66	8	3	9	1	2	1	0
Cost of water	28	58	0	0	0	0	66	35	3	0	1	0	2	6
Inventories, end of period	29	90	65	1	3	2	0	0	1	5	1	0	0	0
Total capital expenditures	1	0	40	1	0	0	30	59	0	0	0	0	2	29
New expenses for buildings	31	49	64	1	0	0	0	0	3	11	1	25	0	0
New machinery	30	48	66	2	1	0	0	0	3	19	1	25	0	0
Value shipments	46	83	44	1	0	0	3	10	3	5	1	0	0	0
Value resales	35	11	63	4	0	0	0	0	0	0	1	2	0	0

**Variability of Nonresponse and Refusal Rates for CPS
October 1981 - December 1982**

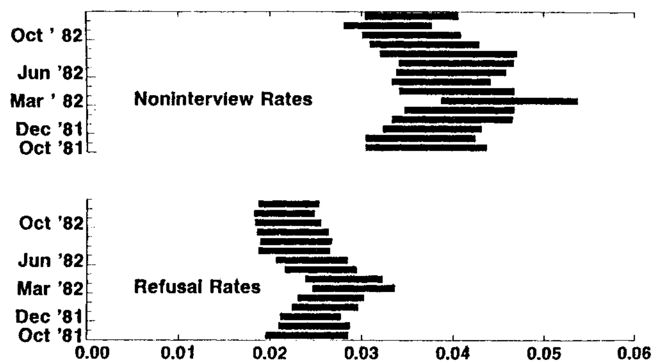

Figure 1

**Response Rates for 88 State and Substate Areas
March 1982**

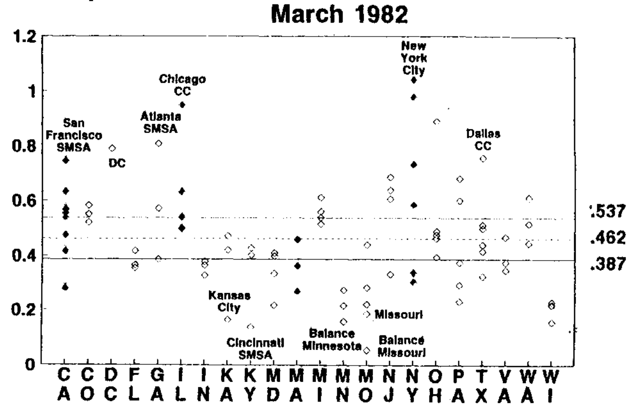

Figure 2

**Chicago City Nonresponse Rates,
October 1981 - December 1982**

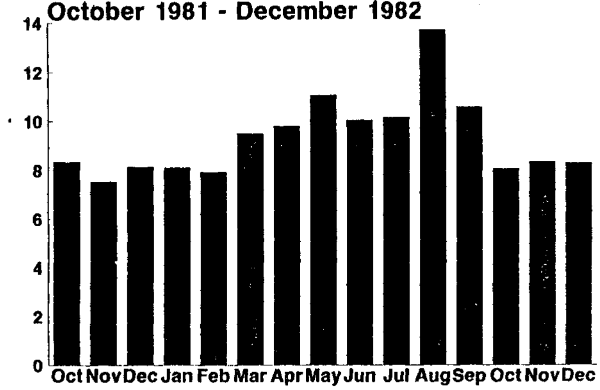

Figure 3