AN OVERVIEW OF CONSUMER EXPENDITURE SURVEY METHODOLOGY

John A. Flueck, Office of Management and Budget [1]
Joseph Waksberg, U.S. Bureau of the Census
Hyman B. Kaitz, U.S. Bureau of Labor Statistics
(Paper presented at 1971 annual ASA Meetings,
Fort Collins, Colorado, August 24, 1971)

Introduction

Consumer expenditures surveys, sometimes termed family budget studies, are one of the older fields of statistical application. Zimmerman [34] and Stigler [32] suggest that partial family budget data, taken from household accounts or roles, date back to the thirteenth century. However, the systematic collection and inspection—surveying—of consumer expenditures appears to have commenced in the late eighteenth century with David Davies and Sir Frederick Morton Eden [2]. By the late nineteenth century Edouard Ducpétiaux, Frederic Le Play, Adolphe Quételet, Ernest Engel, Herman Schwabe, and Carroll Wright (the first Commissioner of the U.S. Bureau of Labor Statistics) had all contributed to this rapidly growing field.

The main objective of the early work was generally on the disclosure of the unfortunate economic and social living conditions, the welfare, of the "working class." Three methodological factors distinguished this work: (1) all families (consumer units) were selected on a purposive basis with some researchers (e.g. Young, Davies, Eden, Le Play) selecting them because of their destitution, (2) all data were collected by interviewers either living with the family (e.g. Le Play) or living in close contact with the consumer unit, and (3) the collection of the data was through a daily, weekly, etc. account or diary.

Over time, as Lamale [16] and others point out, the emphasis of consumer expenditure surveys shifted from the welfare of the poor or working class to the "cost-of-living" of the same groups, and finally to the broader topic of general economic analysis of consumer behavior. Thus one can now list a multiplicity of uses of consumer expenditure data; e.g. economic and social welfare planning, construction of consumption for the national product accounts, weights for price indices, evaluation of present economic policies, and economic or market research.

---

[1] On leave of absence from the Department of Statistics, Temple University.

[2] Davies, an English clergyman, published in 1795 [8] family budget data for 127 agricultural laborer families which he and his fellow-clergy collected in England, Wales, and Scotland. Two years later, Eden [9], in his important three volume work on the English poor, published family budgets for 60 agricultural families and 26 non-agricultural families. Stigler [32] presents the summarized budget data for both studies.

At the same time that the objectives or uses of consumer expenditure data were growing, the methodology used in its collection and analysis were also expanding until there were almost as many methods as there were practitioners. However, the numerous methods (designs) have been classified by Pearl /8/ into four basic categories; (1) the American method, (2) the European method, (3) the panel method, and (4) the split-panel or split-questionnaire method.

The American method is based on interviewer aided annual recall of all expenditures, income, and asset changes for the preceding year with an attempt at their reconciliation. The European method is based on some interviewer collected retrospective information (typically major and regular expenditures) from prior periods, but the main emphasis is on a one-time accounting in a diary for all expenditures during the observation period (typically one to four weeks). The panel method, as the name suggests, relies on successive observations of expenditures over time from a panel of consuming units. Both record-keeping (diary) and personal interviews may be used. Finally, the split-panel method is based on the idea that a series of special surveys (food, travel, income, household furnishings, etc.,) would be coordinated such that all expenditure categories would be covered and a summary of aggregate consumer expenditures could be constructed. Each of these methods has its apparent advantages and disadvantages and both Lamale /18/ and Pearl /19/ have commented on them.

However, no matter which consumer expenditures survey design is followed there are basic methodological problems that must be faced. 3/ The selection of a particular survey design implies particular answers to each of these problems. This fundamental point often appears to receive too little attention.

This paper will attempt to identify a number of basic methodological problems encountered by consumer expenditures surveys, present evidence of their magnitude (response bias and variance) and indicate possible solutions with supporting evidence, if available. Because little meaningful evidence is available on some of these methodological problems, the paper will close with a brief discussion of some suggested research studies which should provide useful guidance in resolving some of these problems. Although the paper is directed at the methodology of consumer expenditures surveys, much of the evidence and discussion is applicable to other types of surveys.

3/ We have chosen to use a general definition of methodology, i.e. a body of methods, rules, and postulates employed by a discipline.

Methodological Problems

Numerous methodological problems arise in any survey, and the attention given to them usually depends on the subject matter, the time available, and their estimated importance 4/ . We have elected to classify the basic methodological problems encountered in consumer expenditures surveys into the following categories: conceptual, response error, sampling and estimation, data collection, data processing, publication, and post evaluation studies. This paper will be essentially devoted to response error problems with limited attention given to some of the other problems.

Conceptual

Every survey design faces the problem of precisely defining the objects (elements of a finite population) and dimensions that it desires to measure. At times the task is relatively easy, so that little attention is given to it. However, even in this situation a mistake at this stage can alter the usefulness of the survey results.

For consumer expenditures surveys, the conceptual task is somewhat complex. First there is the definition problem of what is the unit of observation (i.e. household, family, consuming unit, economic unit, etc.) 5/ . This question includes the problem of how changes in the composition of the observational unit due to marriage, divorce, death, etc. are to be handled.

Second, there is the problem of what observational units are within "scope" for the survey. In short, what elements are in the target population? Only the "urban worker" families which fall within a certain income range? A certain age range? What about their teenage members? These questions and others must be answered in light of the goals of the survey. Generally the survey is focused on "working class" families. Of course this need not be the case. Massey 6/ has reported on a consumer expenditures survey undertaken in England in 1938-39 which focused on "middle class" public officials.

4/ One of the authors 7/ in a paper on the "Essential Steps of Survey Design," has attempted to account for the methodological problems and decisions needed in the design of a survey.

5/ Pearl 8/ points out that although there are a number of different definitions used in the world today, their differences "appear to be more semantic than real." The U.S. Bureau of Labor Statistics definition is: "A group of persons usually living together who pooled their income and drew from a common fund for their major items of expense, or a person living alone or in a household with others but who was financially independent, i.e., his income and expenditures were not pooled " 8/.

A third conceptual problem is what is meant by expenditures? Three alternative definitions are generally cited 5/ which differ on whether the goods and services are measured at the purchase, payment, or consumption stages. The three definitions may produce different results, particularly if loss, waste, "goods" produced at home, and other factors are not controlled for carefully.6/ The option of defining expenditures in terms of sales (by the seller) also appears possible, but most surveys seem to use the "purchases" basis 7/.

A fourth conceptual problem is what "commodities" are to be counted in expenditures? If the answer is all commodities, then gifts, goods produced at home, goods and services in kind, business expenditures, stolen goods, illegal goods (i.e. dope, prostitution, graft, some forms of gambling, etc.), and public services (i.e. use of parks, libraries, public concerts, etc.) must all be measured. Under this definition, the consumer expenditures results should reflect all consumer activities. However, most surveys do not attempt to collect expenditures on the last four types of commodities, and their ability to collect data on the first three types is questionable.

Response Errors

One of the major areas of methodological problems in consumer expenditures surveys is response (or measurement) errors. This is an area which often receives far too little systematic attention.8/ In the design of experiments vernacular, four factors and their interactions (eleven) are eligible to produce response errors. These factors are; the respondent, the enumerator, the instrument, and the environment of the interview. Each of these factors will be examined with some of the interaction terms mentioned where appropriate.

1. Respondent - One of the major problems in consumer expenditures surveys is the acquisition of "accurate" expenditures data from the individual consumer units. Ideally, one would like to monitor and record each and every purchase as it occurs. Practically, this is probably too costly, and thus one must settle for something less than "on-line" recording. The forced compromise is to place the recording burden on a member of the consumer unit (the respondent) and at designated intervals attempt to retrieve the data via a measurement instrument such as a written diary, a schedule, or a tape recorder. The problems attributable to the respondent that arise in this situation are essentially four-fold; will the respondent "recall" the expenditures, will he place them correctly in time, will he submit to the burden (workload) of the data acquisition process, and who is the "optimum" respondent.

6/ Cornfield /6/ presents an interesting example of where an attempt to measure food expenditures via the consumption approach produced an obvious bias in the household reporting.

7/ A notable exception is the U.K. Family Expenditures Survey where the "payment" basis is used for measuring expenditures. See Pearl /9/ for further details.

8/ Two summaries of the response error literature (as it applies to consumer expenditures surveys) which we found most helpful were the Neter and Waksberg /8/ and the Neter /7/ papers.

a. Recall--The "recall" problem can be viewed as having two components; (1) if an expenditure was made will the respondent recall it, and (2) if an expenditure was not made will the respondent erronously recall one. The first case can produce an underreporting problem and the second case an overreporting one. To investigate the magnitude of the recall problem, we examined a number of "record check" studies on products or items that generally appear in consumer expenditures surveys. 9/ Although record checks are not necessarily free of problems, we believe that they represent the best source for estimating the presence and magnitude of the recall problem. 10/

In the food product category the well known Metz /2/ delivered milk expenditure study contained a record check. The study compared respondent (preferably the housewife) answers on the quantity of milk delivered in each of seven previous days with the corresponding dairy records. Table 1 presents the results, and it is seen that the percent of households correctly recalling their milk deliveries decreases as the recall period increases. (74 percent to 59 percent). It is interesting to note that this percentage sequence, or recall decay, does not appear to be linear 11/. Although the bias of the overreporting is larger than the underreporting in each of the seven recall periods, the recall decay of the underreporting is more severe. It increases by 10 percentage points while the overreporting only increases by 5. The net recall bias of 21 percent (overreporting) is suggested by Metz to be due to "social standard" factors. These recall results are rather impressive when one considers that most milk deliveries are fixed orders.

A record check study, involving mainly food items, was conducted by Politz /3/ as a pretest to the Life Magazine sponsored consumer expenditures survey. The study recorded customer expenditures at the check-out counter of six supermarkets, and in the evening of the same day called on the respondent at home and asked that she record the day's purchases in a "sample" diary. The study found that respondents could recall approximately 91 percent of the purchased items and approximately 94 percent of the dollar expenditures. The evidence is that even within a day there is recall decay for food items.

9/ We define a "record check" as a procedure in which a record of an event exists and is checked against the respondents report. As an example, a record of a consumer's credit purchases could be checked against his reported expenditures.

10/ A brief discussion of the general problem of evaluating data, with some comments on potential record check problems, has been presented by Sagen, Dunham, and Simmons /31/.

11/ We have been able to locate only one study, Fisher /13/ that modeled the recall decay. To properly control for recall decay, one must first understand its components and structure, and hence we believe that modeling of response error is an important

A study by Sudman /35/ on the accuracy of consumer panel reporting was also examined. The study compared national estimates of consumer expenditures, derived from diarys from the National Consumer Panel of the Market Research Corporation of America, with manufacturers shipment data (after adjustment for nonhousehold usage) for 55 food, 12 nonfood-grocery and 5 nonfood-nongrocery products. The results showed that 40 of the 55 food products had a negative bias (estimated expenditures less than adjusted shipments) with the bias rising as high as approximately -45 percent, all of the nonfood-grocery products had a negative bias with bias rising to about -35 percent, and all of the nonfood-nongrocery products also had a negative bias which ranged from about -45 percent to -65 percent. It is interesting to note that the grocery products used a weekly diary and the nongrocery products a monthly diary.

An additional 15 products (8, 5, 2) were apparently available toward the end of Sudman's research, and 6 of the 8 food products had a negative bias and all of the 5 nonfood-grocery and the 2 nonfood-nongrocery products also had negative biases. Thus both sets of data are suggestive of substantial recall problems for grocery and household items, even when employing a diary.

In a second product area, consumer durable and semi-durable goods, two studies using record checks were examined. The first study Jaeger and Pennock /16/, compared the respondent's report of the year of purchase of a new refrigerator against the manufacturer's production record. If the two indicated the same year, the respondents report was considered correct. Table 2 presents the results of this record check for the "first collection" (they actually performed the experiment twice on essentially the same set of respondents) of the data. The percent of respondents "correctly" placing the purchase data declines from 51.5 percent to 28 percent" as the households (respondents) reported year of purchase" decreases. 12/ The evidence for recall decay is again present. The surprising result is that only about 50 percent of the respondents who reported their purchase data as within the last five years (0-4) were correct on the year of purchase. One would expect the purchase of a new refrigerator to be a major purchase for most households. We will return to comment on the balance of the table when we discuss the timing component of the recall problem.

The recent Sudman and Ferber research /34/ on recalling durable goods purchases (which you will hopefully hear more about later) was the only other durable goods study we were able to uncover which had some record check results. Table 3 summarizes the record check results for the 39 product classes of their study, and again as the recall period increases the percentage of items correctly reported by the respondent decreases.

12/ Because the year of production of a new refrigerator does not necessarily correspond to the year of purchase, the relationship between the percentages is probably more important than their magnitude.

Also, the underreporting increases by a greater amount than the over-reporting, just as in the Metz milk study (Table 1). However, in contrast to the milk study, the bias of underreporting is now approximately twice that of overreporting. Thus although the evidence suggests that respondent reported durable goods and grocery expenditures are both subject to recall bias, the composition of this bias may be rather complex.

Seven record check studies were located which pertained to the health field Table 4 presents the record check results from the Health Insurance Plan of Greater New York (HIP) study of 1957 /37. A sample of 1400 families was selected from the HIP files, and after interviewing an attempt was made to match their reports for "chronic conditions", with the HIP records. Of the 4,648 chronic conditions inferred from the HIP records, only 44 percent, 28 percent, and 20 percent of the three classes of conditions (Table 4) were reported. 13/ As the table shows the percent of chronic conditions recalled decreases with length of recall for all three classes. It is interesting to note that both the percentage levels and their rate of decay over the three recall periods, are in direct relation to the "clearity" of the probes.

A second record check study in the health field, dealt with the incidence of hospitalization, Cannell /37. A sample of 1,505 persons were interviewed who were known to have been recently discharged from one of twenty-one cooperating hospitals scattered over the continental United States. The total underreporting was 12 percent, but as the recall period increased (week between discharge and the interview, 1-20, 21-40, and 41-53)the percent of underreporting increased (5, 9 and 24 percent). The results also showed that self-respondents were less liable to underreport than other respondents.

Another record check study of incidence of hospitalization /10/ was conducted in the Detroit "urbanized" area in 1961 using records from 18 cooperating hospitals. The results were similar to the previous study. Of the 431 useable interviews conducted with the standard Health Interview Survey questionnaire, hospitalization for the previous year was underreported by 17 percent with the recall loss increasing from 6 percent (1-10 weeks recall) to 32 percent (41-53 weeks recall).

A second record check study on the reporting of chronic medical conditions was conducted on about 2,600 members of the Kaiser Foundation Health Plan, Southern California Region /47. This study was on a "prospective basis" in that the sample members were selected (on a probability basis) and "monitored" by special records for a full year at the end of which they were interviewed. Table 5 presents the record check results and again the evidence is strongly indicative of recall loss and decay. About 55 percent of the overall recorded chronic conditions were reported with 91 percent of the most recent (1-7 days) and only 41 percent of the most distant (365 or more days) recorded conditions
reported. Again, the recall decay appears to follow an exponential relationship over time.

13/ The three classes refer to the use of "checklist" probes in the inter-
view. Class 1 conditions (i.e. diabetes, asthma, high blood pressure, etc.) were clearly probed by the checklists, class 2 conditions had less clear probes, and class 3 conditions had no "obvious" probes.

Two pretests /3/ of X-ray questions for the NCHS Health Interview Survey questionnaire utilized record checks, and although the sample sizes were rather small (92 and 169 completed interviews) the results were suggestive of recall decay.

A record check of the recalling of motor vehicle accident injuries has been reported by Cash and Moss /4/. Table 6 presents the record check results and once again the recall loss and decay is sizable. The percent of respondents recalling the event declined from 97 percent to 73 percent over the 12 months with the percent correctly recalling declined from 92 percent to 61 percent in the same period.

Thus, the recall decay results for the health area appear to be rather similar to those of the other consumer expenditures areas. The longer the period of time since occurrence of the event, the less likely it will be recalled.

Two related areas in which consumer expenditures surveys usually desire to collect consumer unit data are income and assets. In the area of income, three record check studies were examined. David /7/ reported on the interviewing of 46 families who had received public assistance for at least 12 months prior to the interview. He found that respondents, on average, underreported by 18 percent the amount of public assistance they received (average of $2,334), and that 26 percent underreported by more than 30 percent. A Census Bureau study /9/, using a larger sample of people on public assistance (411), also performed a record check on reported income. The study found that recipients underreported the magnitude of assistance by about 27 percent with underreporting occurring about twice as often as overreporting. Two additional findings are worth noting, (1) the median amount of underreporting decreased as the amount of assistance increased, and (2) the tendency to underreport increased with the number of months on assistance. 14/

The Census Bureau, as part of its 1960 Census Evaluation Program, performed a record check between reported income on the Census and income reported to Internal Revenue Service /5/. The results were rather good for wages and salaries; net underreporting to census of about 3 percent with some difficulties in the tails of the distribution, but rather substantial overreporting of self-employment income (+43 percent) and income other than earnings (+48 percent).

The implication for consumer expenditures surveys of these income results is that certain types of income may be difficult to "accurately" collect by the standard survey techniques. New approaches may be needed for certain types of income.

Three record check studies of asset reporting on household interviews were examined. The first study, Lansing, Ginsburg, and Braaten /19/,

---

14/ Purposive misreporting could be viewed as a recall problem. The "accurate recall" of an event would then have as one of its arguments the "type" (socially desirable, illegal, etc.) of event.

involved three small-scale (samples of about 90 interviews) record checks of saving account balances. The findings were that about 30 percent were nonrespondents, underreporting was dominant, and recall decay was in evidence, but trival.

A larger-scale record check study of savings account balances has been reported by Ferber et al /13/. The results were again, (1) nonreporting is the principal source of error, about 44 percent, and (2) nonreporters typically had larger balances. They found the mean balance was underreported by about 46 percent. Ferber et al also performed a record check study of common stock holdings /11/ and the results were similar to the savings accounts study.

Thus, the asset reporting picture is not very bright for consumer expenditures surveys. However, one might argue that respondents should be more willing to report changes in assets, than levels of assets, but we are rather doubtful. Accurate asset reporting on household interviews, as Feber and others have long ago pointed out, is a tricky unresolved problem.

A second technique, the comparison of different recall periods, has been used in a number of studies to investigate errors of recall. 15/ The Neter and Waksberg /28/ and Neter /21/ papers summarize eight of these studies (Ghosh, Lewis, Mahalanobis and Sen, Mahalanobis and Sen, Cole and Utting, Goldberg, Kemsley, and Neter and Waksberg) pertaining to consumer expenditures. In six of the eight studies, covering numerous food and nonfood products, the results showed that longer recall periods resulted in less purchases. In the seventh study, only about one-half of the products indicated less purchases for the longer recall period when compared to the weekly recall period. The eighth study, a small scale preliminary study by Ghosh /13/, also indicated less purchases (9 food items) for the longer recall periods (daily, weekly, annual), but estimates of household food purchases based on shopping activity in the same three villages were closest to the annual recall values. For 6 of the 9 items, the daily recall values were above the estimated purchase values.

Although there may be a timing bias present in this study, we feel that it is the daily purchase estimation process which is suspect. For Ghosh also provides daily purchase estimates based on three other villages which are "nearer the city area" but were "not accompanied by any household inquiry". He suggests that these results also "reveal close similarity in the estimates" and indicate "a general uniformity in the volume of purchase by villagers.....". However, if one compares the mean estimated daily purchases of these three (no household interview) villages with the mean estimated daily purchases of the original (household interview) three villages, 8 out of the 9 items have substantially higher values for the "no household interview" villages. Furthermore, if one compares these "no household interview" estimated daily purchases with the mean daily purchases reported by the households in this first set of villages, 7 of the 9 items now have reported mean purchases that are smaller than the estimated mean purchases. Under this comparison, the observed decline in reported mean purchases with increasing recall periods accords well with other recall results.

_____

15/ This technique does not allow one to directly compute the magnitude of the recall error. However, the relative recall error and decay (possibly confounded with a timing error) can be observed.

The previously mentioned Sudman and Ferber study contained data on varying periods of recall (1-12 months) and it too found substantial evidence of a recall error and decay over time.

Thus the results from the record checks and recall period comparisons are both strongly suggestive of a recall error which increases (possibly exponentially) with longer recall periods for virtually all consumer expenditure items. The important question is the functional form of the recall error for the various commodities in a consumer expenditures survey, and here the evidence is not very clear. Although it is tempting to speculate on this functional form, we will take the more cautious course and refrain (by a vote of 2 to 1) from so doing at this time.

b. Timing--The timing response error problem (sometimes termed telescoping) is concerned with the respondents "dating" of the event. The options are to shift the event forward in time, or backward in time, or time it correctly. Interest is often placed on the net timing error which is the difference between the forward and backward shifting. As before, we again place primary emphasis on the measurement of timing error through record checks.

The Jaeger and Pennock new refrigerator record check study /16/ also presents some timing results. Returning to table 2, note that about 40 percent of the respondents tended to shift the purchase date forward in time for both the 0-4 and 5-9 year "recall periods", as opposed to 9 percent and 17 percent shifting the event backward. For the longest recall period, "10 years or more", the results are reversed, 52 percent shift the data backward in time and 20 percent shift it forward. Even though the sample sizes are rather small, this reversal is rather surprising. Aggregating over all three "recall" categories, the net timing error is forward and about 17 percent.

The Detroit hospitalization record check study mentioned earlier also contained some timing results. Of those who recalled their hospitalization, 77 percent reported their month of discharge correctly, 14 percent shifted it forward, and 9 percent shifted it backward. Again there is a net forward shifting (positive timing error) of the event.

Lastly, the previously mentioned Cash and Moss /4/ record check study (motor vehicle accident injuries) also contains some timing results. Table 6 presents the evidence, but a word of caution is needed. The timing results for the " <3 months" and "9-12 months" recall periods should probably be given little attention because the interviewers bounded the recall period with two dates. The remaining two recall periods both show a higher percentage of forward shifting than backward shifting (by a magnitude of 4) with the forward shifting increasing with the length of the recall period.

Moving from studies that employed record checks to studies that relied on comparing different recall periods, the Neter and Waksberg study /18/, through its use of bounded recall, presents substantial evidence on timing bias. This study, Survey of Residential Alterations and Repair (SORAR) mainly utilized personal interviews of approximately 10,400 households, interviewed them at most four times, and chiefly used bounded and unbounded one and three months recall periods.

The most impressive finding was the evidence of forward shifting of both jobs and expenditures. Comparing unbounded one month recall with bounded one month recall, they found a net forward shifting of 40 percent for reporting of jobs and 55 percent for reporting of total expenditures. These timing errors were far beyond any sampling variability, and they appeared to hold over all types of jobs (i.e., do-it-your-self, contract, paid help, etc.) and over all sizes of jobs. There was also evidence that the net timing error was greater for larger jobs than for smaller ones. The net error for reported jobs was about 56 percent for jobs over $100 and about 29 percent for jobs under $10. The corresponding figures for total cost of the job (expenditures) were 54 percent and 35 percent. Neter and Waksberg also compared unbounded and bounded three month recall and although the results were not quite as strong there was again evidence of a positive timing error of about 4 percent for jobs and 19 percent for expenditures. The already mentioned Neter paper /17/ contains more discussion of these and other results.

In summary, with few exceptions, we find strong evidence of a forward shifting of events (a positive time bias) when no controls (such as bounding) are used. This has strong implications not only for one time consumer expenditures surveys, but also for continuing expenditures surveys.

c. Workload--The respondent workload problem evolves from the question of how much work (burden) the respondent will submit to during the data acquisition process. Clearly, the more unremenerated work asked of the respondent, the smaller the chances of full cooperation. The quality of the elicitted data is, no doubt, also tied to the workload.16/

However when one searches for evidence in the consumer expenditures field-- or any field-- in order to examine the above beliefs, the pickings appear slim. The Neter and Waksberg study presents some evidence that suggests when the reporting period is lengthened, the quality of the data deteriorates. Neter /17/ summarizes this and a few other results.

If one is willing to use nonresponse rates as a proxy measure for work-load, then a consumer expenditures diary study by Sudman and Feber /33/ is of interest. Although we will hopefully hear more from the authors about this study, two points standout with respect to workload. First, the response (cooperation) rates for diary households were directly related to the number of weeks a household was requested to keep a diary. As an example, of those respondents requested to keep a diary one week, 90 percent did so, Those requested for two weeks, 87.5 percent did so. For three weeks, 85.7 percent, and for four weeks, 84.7 percent. Second, the diary response rates were noticably higher when a gift was offerred the household than when it was not. As an example, for those households that kept at least one diary the response rate for those who were offerred a "large stationary holder" was 90 percent versus 77 percent for the "no gift" households. In short the message for consumer expenditures seems clear. Serious consideration should be given to remunerating respondents when the workload is large.

16/ Certainly both of these points are of direct interest to all "surveyors." The Federal Reports Act of 1942 is one example of federal governments concern in this area.

d. Optimum Respondent--Information on the selection of the optimum
    respondent within a household is incomplete. There is evidence
    that in some fields (e.g. health) that self-respondents report
    substantially higher incidences than an undesignated household
    respondents /3/ /3/ . In the Current Population Survey studies
    /17/ there is also some evidence of respondent effect in reporting
    labor force data, particularly in the unemployed female category.
    For consumer expenditures, various results have been reported.
    In a small scale study, Ferber   has found substantial differences
    arising from different respondent selection rules. On the other
    hand, Neter-Waksberg /8/ found little indication of differences
    among, heads, wives, head and wife, and unspecified respondents.
    Hence, the question of an optimum respondent is very much open.
    It seems intuitively plausible that certain respondents may be
    optimum for one type of expenditure (i.e. lunches eaten out, auto
    repairs, etc.) and other consumer unit members optimum for others.

2. Enumerator - Although there is limited information on the variability
    in enumerator skills and performances, available evidence from both
    consumer expenditure and other research suggests that enumerator effects
    can be important. The Metz study /3/ showed a significant difference
    among enumerators with respect to the bias in reporting delivered milk
    quantities. In particular, the poorest enumerator's work had three
    times as large a bias as the next extreme case. Neter /1/ summarizes
    this and other studies showing similar, but perhaps not as dramatic,
    enumerator effects. Major enumerator effects on population census
    data have been reported by both the U.S. Bureau of the Census /3/ and
    the Dominion Bureau of Statistics /10/.

    The implication of these results should be carefully considered in
    designing consumer expenditure surveys. In particular, relying upon the
    work of a single or only a few enumerators raises potentially serious
    problems with local area data. The opportunity for off-setting
    enumerator effects, through pooling of considerable data, is not present.

    A second issue concerns the enumerator effects in one-time surveys.
    Analyses of interviewer performance of a number of Census Bureau surveys
    indicate that a substantial enumerator learning experience is needed.
    For example, Figure 1 shows that it required over a year (12 monthly
    interviews) for enumerators' error rates to be reduced to a relatively
    low and stable level in the Current Population Survey. The message for
    consumer expenditure surveys appears to be that the more interviewing
    opportunities the enumerator has over a period of time the greater the
    chance of producing high quality data. It is obvious that a "continuing"
    consumer expenditure survey program would considerably benefit from such
    a "learning" process.

3. Measurement Instrument - The problem of meaningfully quantifying
    the desired information -- measuring -- is an all important one.
    For us, measuring entails not only the instrument utilized
    (schedule, diary, voice recording, etc.) but also the process
    (self- enumeration, interviewer physically present, telephone, etc.)
    Although there has been considerable interest and research in
    the general measuring process, the results are rather mixed.
    Neter /21/ summarizes much of these results for consumer expenditures
    surveys. In summarizing the "global" versus detailed question
    research, he suggests that "increasing the probing, detailed nature
    of the questionnaire or changing the setting of the questionnaire
    does affect reports up to a point, but past this point further
    changes in the questionnaire have little effect". We are inclined
    to agree with Neter in this generalization, but with one important
    qualification. It is conceivable that probing and the use of
    detailed questions when pushed beyond "the point" will have a
    negative effect on the respondent and the data. It seems to us
    that rather little is really known about the benefits and lia-
    bilities of this important tool.

    The question of whether a schedule or diary is the best procedure
    for collecting consumer expenditures has received attention. However,
    it seems to us that much of the research has not properly controlled
    for other factors. Certainly the problem of conditioning is present
    in comparing the schedule with the diary, and it must be meaningfully
    taken into account. In fact conditioning is probably an inter-
    action effect between the instrument, respondent, workload, and
    possibly the enumerator. Neter /21/ summarizes much of the material
    on conditioning, and it seems to us that the proper measuring of
    conditioning will probably not be an easy task.

    Another subject of interest in the measurement instrument area is
    the use of the telephone. Certainly the telephone has potential
    for producing virtual on-line-recording of expenditures on a daily
    basis. Like any good thing, it also contains its own set of problems.
    One of these problems appears to be its proclivity to produce lower
    reported expenditures, Neter and Waksberg /23/. Perhaps new inter-
    viewing techniques are needed to adequately use the telephone in
    consumer expenditures surveys. In fact, with the acknowledged problems
    in the accurate reporting (acquisition) of some types of income,
    assets, socially undesirable activities, etc., it seems to us that
    experimentation in new measurement instruments is needed.

4. Environment - We were unable to locate any consumer expenditure studies which examined the question of the optimum interview environment. In theory, many alternative environments are available, but in practice it seems that only the household interview setting is used. However, even this environment has some variations. Certainly the time of day of the interview changes some aspects of the environment. Also the number of respondents present at the interview may have an effect on responses. It may even be that the evidence on underreporting of expenditures via telephone is more of an environment than an instrument problem. It may be that the environment needed for a successful telephone interview is considerably different than a household interview.

It seems to us that the technology is already available for virtual on-line-recording of consumer purchases. Whether it will ever come to this is another matter. However, we do believe that some experimentation is needed on interview environment for consumer expenditures surveys, particularly in light of their demonstrated above average workload and the apparent sensitivity of some of the data.

Sampling and Estimation

As far as sampling is concerned, we do not see any problems that are unique to consumer expenditure surveys. Standard sampling techniques are applicable.

However, there are a number of estimation problems that should be investigated. The usual estimation decisions inherent in any sample survey are present, such as whether to use unbiased estimates, ratio or regression estimates, etc. There are also a number of more specialized estimation problems. For example, an important question that can be raised is whether the same estimation techniques should be used in deriving both CPI weights and family budget data. This decision affects treatment of in-movers, out-movers, births, deaths of consumer units, as well as the method of combining diary and interview results.

Another issue involves the method of producing data for local areas as opposed to national or regional statistics. There are two views that can be taken. The first puts the prime emphasis on obtaining the best possible national figures, with local data being a by-product. This is the approach taken in the Current Population Survey. The second gives more emphasis to the production of local area statistics, and the resulting oversampling in these areas weakens the national statistics by increasing their variances. This approach has been used by BLS in the national consumer expenditure surveys. With the sample sizes that have been used in the past, there is some question whether the local area statistics are of sufficient reliability to justify the weakening of the national figures.

An alternative to oversampling in a local area is to use pooled data for a set of similar areas in order to increase the sample size. This estimation technique is not new, and has recently been referred to as "synthetic estimation" [25]. Recent theoretical work by Waksberg [35] gives the conditions under which synthetic estimators are more reliable (in the mean square error sense) than estimators based entirely on the local area data. The use of this estimation technique in Consumer Expenditure Surveys should be explored.

Post Evaluation

All important surveys should have solid post evaluation plans. The importance of such planning is twofold; (1) it forces you to think constructively about the numerous survey problems and how you might encourage the survey to provide you with evidence on them, and (2) it supplies a useful foundation for improving the performance of the survey at its next running. The Evaluation and Research program of the Census Bureau is an example of solid post evaluation.

For consumer expenditures surveys numerous conponents are available for establishing a strong evaluation program. Record checks of a subsample of respondents are possible. Numerous internal checks are also available and have been used by BLS in the past. Comparisons with National accounts, other expenditure item surveys, and other government records should be encouraged. Certainly the post evaluation plans should include a program to measure the effects of some of the previously mentioned response errors and other survey problems. Finally, one of the best post evaluation studies is to make the entire data base available to researchers and users. Extensive use of a data base may in the final analysis be the best source of evaluation.

Suggested Research Studies

One point that comes through strongly after reviewing the numerous studies in consumer expenditures surveys methodology is the pressing need for additional, well designed and controlled, experiments. Many of the present response error results appear to be "after-thoughts" or auxiliary findings from a main study with the result that it is difficult to factor out the effects. Certainly, the time has come to put a major emphasis on a unified approach to response error studies in the field of consumer expenditures surveys. These studies should include both "modeling" and the estimates of population parameters of the response error models.

It seems to us that three approaches are available for increasing the understanding, modeling, and controlling of response errors. First, some surveys should be designed solely for experimentation purposes relying on the theory and concepts of the Design of Experiments. Second, it is possible to do some experimentation within one-time surveys, particularly if multiple visits are to be used. Record checks, randomization over interviewers, bounded compared with unbounded recall, etc., are all possible. The Neter and Waksberg /16/ study of SORAR is an excellent example of this technique. Perhaps all large surveys should automatically have attached to their funding requests a percentage for experimentation. This percentage may not have to be very large, for the cost of experimenting with a subset of respondents from a terminating survey should be relatively low. Finally, it is possible to experiment within continuing surveys, using such techniques as Evolutionary Operation (EVOP).

Box /17/ developed the EVOP technique so that one can search within an on-going industrial process, without disturbing its production quality, for that composition of inputs which will maximize output. It seems to us that this technique has great potential for response error investigation within on-going (continuing) surveys. The goal will be to find that composition of survey "inputs" such that you move toward the maximizing of accurate responses and thus the minimizing of response error. This concept has been profitably used in the Current Population Survey. This technique also may have applicability in other continuing surveys.

Having taken a strong stand on the need for increasing our knowledge of response error, we propose that a continuing research program on this subject be instituted including the incorporation of some well designed response error experiments within the forthcoming National Consumer Expenditure Survey (CEX) program. The following six experiments are of the type that we would like to see considered.

1. **Record Check**

   Reliability and validity checks should be made on consumer expenditure data. Matching store records with reported expenditures will provide useful information on the quality of the data.

   We propose that in a number of PSU's, arrangements be made for securing charge account records from major stores and suppliers for designated sample households. A check will be made whether the indicated purchases were reported in the family's expenditures report.

2. **Balancing of Income and Expenditures as a Probing Tool**

   To go from an annual recall one-visit survey to a quarterly survey, we have sacrificed the balancing and probing procedure. In this method, total income for the year was compared with reported expenditures and reconciliation of differences was attempted with the respondent. In the interest of determining the importance of the sacrifice, an experiment could be conducted to measure the potential effect of balancing as a probing device. Since the present plans for the national consumer expenditure survey are to collect approximately 95 percent of all expenditures on the quarterly schedule, a balancing operation cannot be performed until the additional 5 percent of the expenditures have been obtained from the individual respondents.

   We propose that for a subset of households a sixth interview be conducted which would collect the remaining expenditures and thus permit a balancing and probing procedure. If the resulting changes are minor, then the balancing operation is of questionable use. However, if there are major charges, additional work will be needed to determine if the charges actually improve the expenditures records.

3. **Study of Alternates to Diary Approach**

   Past experiences with diaries suggest that response rates on diaries, and the quality of the data, is at times rather low. For items which have small week-to-week variability (e.g. food, gasoline purchases, lunches, certain household supplies) we propose an experiment to study the effectiveness of alternative methods of collecting expenditure data. These alternatives would be applied for two weeks to subsamples of the diary households following the diary collection with one exception.

(a) In the first subsample, two one-week recalls, using a personal visit, would be attempted covering expenditures for the selected above-mentioned items.

(b) Another subsample would also utilize two one-week recalls, but employ telephone interviews.

(c) In a third group, daily telephone interviews would be conducted for the entire two week period.

(d) In order to factor out possible conditioning effects, a fourth subsample should be selected from nondiary households and a bounding performed, followed by daily telephone interviewing for the two-week period.

The procedures could be evaluated by comparing aggregate levels with the diaries for the same households, under the assumption that more is better. The fourth subsample could be compared with the third.

4. Split Questionnaire Technique

Information is needed on the trade-off between the reporting burden and quality of data. We propose that a split-questionnaire design be used to study this problem.

We suggest that all expenditures be split into a number of categories, with an independent subsample of households used to collect data for each category. Comparisons with national CEX totals should indicate the cost, in quality, of requiring complete expenditures for each respondent.

5. Incentives for Promoting Cooperation

This survey requires a much longer period of cooperation by the respondent than the normal government survey. The result has typically been a high nonresponse rate and most likely lower quality data.

Accordingly we propose that incentive devices be tried to improve response and quality. Specifically, a varying schedule of monetary payments and gifts should be tried.

6. Increasing the Number of Respondents

The question arises whether increasing the number of respondents in a consumer unit results is better. Two experiments are proposed.

(a) The quarterly schedule interview to be conducted with both husband and wife present.

(b) For a set of diary households a review of the diary would be conducted with all teenager and other adult members of the consumer unit. Additional expenditures would be recorded and analyzed.

Clearly the above list does not exhaust the set of useful response
error experiments. Considering the potential benefits from the above
suggested research, we feel that an immediate effort should be made
to initiate many of these studies, with definite plans for a continuing
program to study the rest. Certainly, some of the above experiments
could be easily integrated into the present plans for the national CEX.

Lastly, the existence of numerous methodological problems in the
consumer expenditures survey field should be clear. We hope that the
presented evidence and discussion is helpful in providing solutions to
these important problems.

TABLE 1. Results from a record check of delivered milk
expenditures in a Central New York Community.

Recall Period	Number of Households	Percent of Households Reporting Quantity		
		Correct	Under	Over
1 day	1893	74%	8%	18%
2	"	76	8	16
3	"	72	8	20
4	"	73	10	17
5	"	69	12	19
6	"	65	12	23
7	"	59	18	23
mean Bias (percent)		0	22	43

Source: /22/

TABLE 2. Results from a record check of the year
of purchase of a new refrigerator, two cities.

Household Report on Purchase Date	Number of Households	Percent Placing Date of Purchase		
		Correctly	Earlier	Later
0-4 years ago	68	51.5%	8.8%	39.7%
5-9	47	42.6	17.0	40.4
10 or more	25	28.0	52.0	20.0
Total	140			
Mean Bias (percent)		0	19.3	36.4

Source: /16/

TABLE 3. Results from a record check of consumer durable goods purchased in three Illinois towns.

Recall Period	Number of Items verified	Percent of Items		
		Correctly reported	Not reported (under)	Not in store record (over)
3	325	14.8%	50.1%	23.4%
6	561	11.6	48.5	30.7
12	822	10.7	56.0	26.4

Source: /34/

TABLE 4. Results of a record check of inferred chronic conditions using the H.I.P. New York City data.

Recall Period	Number of Conditions Inferred from Record			Percent of Conditions Reported on Household Interview		
	Class 1	Class 2	Class 3	Class 1	Class 2	Class 3
0 - 1/2 months	246	87	124	67.9	50.6	41.9
1/2 - 4	714	413	602	49.3	34.1	22.1
4 $\leq$	912	731	819	33.7	21.2	15.9

Source: /23/

TABLE 5.   Results from a record check of chronic medical conditions
using Kaiser Health Plan Southern California data.

| Recall Period	Number of Recorded Conditions	Percent of Conditions Reported in Interview
1-7 days	116	91.4
8-14	218	72.5
15-28	440	76.1
29-56	683	58.4
57-84	574	63.1
85-112	513	57.7
113-140	476	54.8
141-168	355	53.8
169-224	372	43.5
225-280	1,232	48.2
281-364	1,078	42.4
365 or more	71	40.8
Total	6,128	

Source: /24/

TABLE 6.   Results from a record check of motor vehicle accident injuries
using data from three North Carolina counties.

Recall Period	Number of Completed Interviews	Percent Recalling Event	Percent Placing Date of Event		
			Correctly	Earlier	Later
less than 3 months	119	97%	92%	5%	0%
3-6 months	209	90	79	2	8
6-9 months	119	86	70	3	12
9-12 months	143	73	61	0	12
	590				

Source: /4/

Figure 1. MEAN ERROR AND NONINTERVIEW RATES FOR CPS ENUMERATORS

(For enumerators trained in '62
and '63. Moving averages with
period = 3)

CUMULATIVE MONTHS WORKED
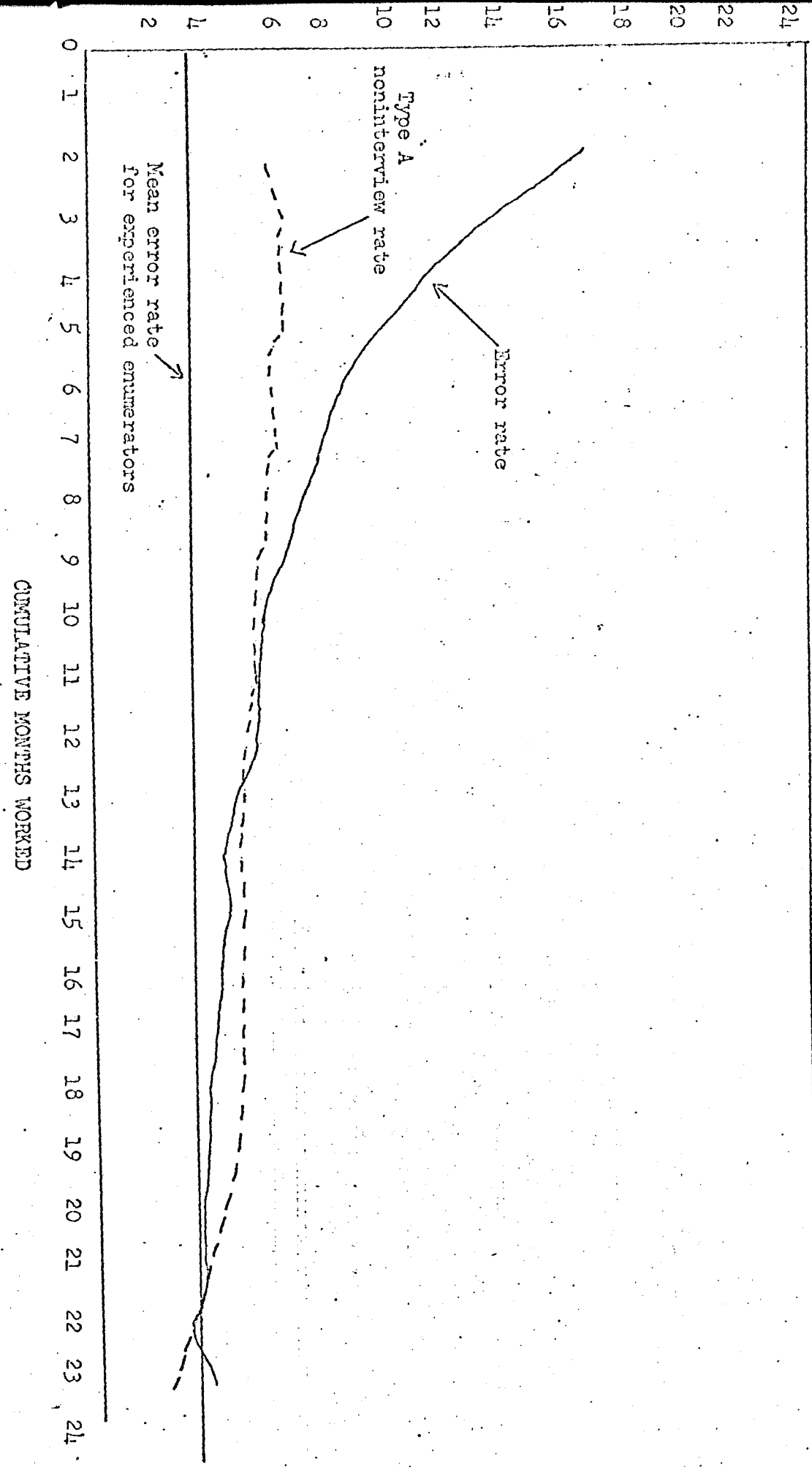
References

1. Box, G.P. and Draper, N., *Evolutionary Operation*, Wiley, New York, 1969.

2. Bureau of Labor Statistics, *Consumer Expenditures and Income*, Total U.S., Urban and Rural, 1960-61. BLS Report No. 237-93, February 1965.

3. Cannell, Charles, "Reporting of Hospitalization in the Health Interview Survey", *Vital and Health Statistics*, NCHS, Series 2, No. 6, Public Health Service, 1965.

4. Cash, W.S. and Moss, A.J., "Methodology Study for Determining the Optimum Recall Period for the Reporting of Motor Vehicle Accidental Injuries." *1969 Social Statistics Section Proceedings*, ASA 1969 Annual Meeting, 1969.

5. Census Bureau, *Evaluation and Research Program of the U.S. Census 1960: Record Check Study of Accuracy of Income Reporting*, Series ER 60, No. 8, Washington, D.C., 1970.

6. Cornfield, Jerome, "On Certain Biases in Samples of Human Population", *Journal of the American Statistical Association*, Vol. 37(1942), pp. 63-68.

7. David, Martin, "The Validity of Income Reported By a Sample of Families Who Received Welfare Assistance During 1959", *JASA*, Vol. 57 (1962), pp. 680-685.

8. Davies, David, *The Case of Labourers in Husbandry*, Bath, 1795.

9. Eden, Sir Frederick M., *The State of the Poor*, London, 1797.

10. Fellegi, I.P., "Response Variance and its Estimation", *Journal of American Statistical Association*, Vol. 59 (1964), pp.1016-1041.

11. Ferber, R., Forsythe, J. Guthrie, H.W., and Maynes, E.S. "Validation of Consumer Financial Characteristics: Common Stock", *JASA*, Vol. 64 (1969), pp. 415-432.

12. Ferber, F., Forsythe, J., Guthrie, H.W. and Maynes, E.S., "Validation of a National Survey of Consumer Financial Characteristics: Savings Accounts", *The Review of Economics and Statistics*, Vol. 51 (1969), pp. 436-444.

13. Fisher, Gordon, "A Discriminant Analysis of Reporting Errors in Health Interviews" *JRSS* (Series C, Applied Statistics), Vol. 11 (1962), pp. 148-163.

14. Flueck, John A., "The Essential Steps of Sample Survey Design", Unpublished memorandum, Office of Management and Budget, Executive Office of the President, Washington, D.C., 1971.

15. Ghosh, A. "A Note on the Accuracy of Family Budget Data with Reference to the Period of Recall", Calcutta Statistical Association Bulletin, Vol. 5 (1954), pp. 16-23.

16. Jaeger, Carol and Pennock, Jean, "A Note on Accuracy and Consistency of Response in Household Surveys", Unpublished Memorandum, Agricultural Research Service, Department of Agriculture, 1962.

17. Jones, Charles and Aquilino, Robert, "Method Test Phase III: Second Report on the Accuracy of Retrospective Interviewing and Effect of Nonself Response on Labor Force Status", Unpublished memorandum, Bureau of the Census, Department of Commerce, January 29, 1970.

18. Lamale, Helen H., Study of Consumer Expenditures Incomes and Savings, Methodology of the Survey of Consumer Expenditures in 1950, Wharton School of Finance and Commerce, University of Pennsylvania, 1959.

19. Lansing, John B., Ginsburg, Gerald P., and Braaten, Kaisa, An Investigation of Response Error, University of Illinois Bureau of Economic and Business Research. Studies in Consumer Savings, No. 2, June 1961.

20. Livingston, Rockwell, "Evaluation of the Reporting of Public Assistance Income in the Special Census of Dane County, Wisconsin", Paper Presented at the 9th Annual Workshop of Public Welfare Research and Statistics Personnel, New Orleans, August 17, 1969.

21. Massey, Philip, "The Expenditure of 1,360 British Middle-Class Households in 1939-39", Journal of the Royal Statistical Society, Vol. 55 (1942), Part III, pp. 159-180.

22. Metz, Joseph F., "Accuracy of Response Obtained in a Milk Consumption Study, Paper No. 5, Methods of Research in Marketing, Cornell University College of Agriculture, Ithaca, N.Y., July, 1956.

23. National Center for Health Statistics, "Health Interview Responses Compared with Medical Records, "Vital and Health Statistics, PHS Pub. No. 1000, Series 2, No. 7, Public Health Service, Washington, D.C., July 1965.

24. National Center for Health Statistics, "Interview Data on Chronic Conditions Compared with Information Derived from Medical Records", Vital and Health Statistics, PHS Pub. No. 1000 - Series 2, No. 23, Public Health Service, Washington, D.C. May 1967.

25. National Center for Health Statistics, Synthetic State Estimates of Disability, PHS Pub. No. 1759, Washington, D.C., 1968

26. National Center for Health Statistics, "Volume of X-Ray Visits", Health Statistics, PHS Pub. No. 584-B38, Public Health Service, 1962.

27. Neter, John, "Measurement Errors in Reports of Consumer Expenditures", Journal of Marketing Research, Vol. VII (1970), pp. 11-25.

28. Neter, John and Waksberg, Joseph, Response Errors in Collection of Expenditures Data by Household Interviews: An Experimental Study, Technical Paper No. 11, Bureau of the Census, Department of Commerce, 1965.

29. Pearl, Robert B., Methodology of Consumer Expenditures Surveys, U.S. Department of Commerce, Bureau of the Census, Working Paper No. 27, 1968.

30. Politz, Alfred, "Description of Operational Design and Procedures", Life Study of Consumer Expenditures, Vol. 4, Time Inc. 1958.

31. Sagen, O.K., Durham, R.E., and Simmons, W.E., "Health Statistics from Record Sources and Household Interviews Compared", Proceeding of the Social Statistics Section, ASA Annual Meeting, Washington, D.C., 1959.

32. Stigler, George J., "The Early History of Empirical Studies of Consumer Behavior," The Journal of Political Economy, Volume 62 (1954), pp. 95-113.

33. Sudman, Seymour and Ferber, Robert, Experiment s in Obtaining Consumer Expenditures By Diary Methods, University of Illinois, report to BLS, Department of Labor, April 1970.

34. Sudman, S. and Ferber, R., Experiments in Obtaining Expenditures of Durable Goods by Recall Procedures, University of Illinois, report to BLS, Department of Labor, August 1970.

35. Sudman, Seymour, "On the Accuracy of Recording of Consumer Panels: I", Journal of Marketing Research, Vol. 1(1964)pp. 14-20.

36. U.S. Bureau of the Census, <u>Evaluation and Research Program of the U.S.</u>, Census 1960: Effects of Interviewers and Crew Leaders, Series ER 60 No. 7, Washington, D.C. 1968.

37. United Nations, <u>Recommendations Concerning Family Budget Inquiries in Europe</u>, Conference of European Statisticians/W G. 17/6, May 24, 1962.

38. Waksberg, Joseph, "Analysis of Synthetic Estimates" unpublished memorandum, Bureau of the Census, Department of Commerce, December 10, 1970.

39. Zimmerman, Carle C., <u>Consumption and Standards of Living</u>, D. Van Nostrand Co., Inc., New York, 1936.

40. National Center for Health Statistics, "Comparison of Hospitalization Reporting,"<u>Vital and Health Statistics</u>, PHS Pub. No. 1000, Series 2, No. 8, Public Health Service, Washington, D.C., July, 1965.