A TEST OF DATA COLLECTION METHODOLOGIES: THE METHODS TEST

Charles D. Cowan, Anthony M. Roman, Kirk M. Wolter, and Henry F. Woltman, U.S. Bureau of the Census

1. Introduction

The Methods Test Panel (MTP) is a survey research vehicle designed by the Bureau of the Census to test alternative methodologies and concepts used in the Current Population Survey (CPS). The MTP is an attempt to improve the quality, reliability, and utility of CPS data, and is intended to provide a way of testing and evaluating recommendations from the National Commission on Employment and Unemployment Statistics (NCEUS). Although the project addresses issues that directly relate to labor force data collected in the CPS, the methodological findings should have application to a broader class of household surveys.

During the mid 1960's the Bureau conducted similar tests in connection with recommendations made by the Gordon Committee (1962), a Presidential committee appointed to appraise the labor force statistics available at the time. Results from those tests are described by Waksberg and Pearl (1965). The first series of experiments in this ancestral MTP were directed toward modifications in questionnaire design and content and toward interviewing procedures. These experiments were conducted over a span of 21 months, and resulted in improvements in the measurement of hours worked and the reporting of self-employed status. Subsequently, a second series of experiments were initiated for the purpose of testing self-response procedures. No significant differences between the self-response and customary CPS procedures were found in the estimation of unemployment.

The current MTP has also been organized into two or three separate phases or groups of experiments. The first phase has, as of this presentation, been in the field for 16 months and is being used to test alternative data collection methodologies. The customary CPS instrument is used in this phase. A second test is being prepared to replace the first in December 1979; it will test alternative question wordings suggested by the Bureau of Labor Statistics and the NCEUS. Funding for the MTP is slated for 4 years, so following the second test a third may be instituted which would deal with final recommendations from the Commission or other topics where testing seems useful.

One of the main goals of the MTP project is to provide information which would be useful in directing the next major redesign of CPS. The next comprehensive redesign is scheduled for the early 1980's.

This paper deals solely with the design of the first phase tests, and a partial analysis of the data collected between May 1978 and November 1978. Section 2 describes certain potential nonsampling problems in the CPS and the subsequent selection of experimental treatments used to study these problems. The sample and experimental designs for MTP are discussed in Section 3, while Section 4 presents some examples of the kinds of analyses that are being used for these data. Section 5 closes the paper with a general summary.

2. Choice of Experimental Treatments

2.1 Some Potential Problems in the Current Population Survey

There are a number of potential nonsampling problems in the CPS, and although some research has been conducted on these problems, there is a limit on large-scale experimentation in the CPS because of the importance of the data and the fear that experimentation may disrupt the various labor force time series. Hence the necessity of a separate research panel like the MTP.

We cite three example problems. One important problem is rotation group bias. See, e.g., Bailar (1975). This bias arises because of the rotating panel structure of CPS, meaning that households are interviewed repeatedly according to a specific pattern, with some households retired from sample each month, and new households being rotated into sample to replace the retirees. A 4-8-4 rotation pattern is used, where housing units are interviewed 4 successive months, retired for 8 months, then rotated back into sample for 4 final months. Each month in sample (or rotation group) is itself a nationally representative probability sample, each of which should have the same expected value, and, as a result, the number of significant differences among the estimates prepared from the eight groups should be within the range expected by chance. The fact that the number of significant differences exceeds what would be expected by chance has caused the Bureau great concern, and the underlying mechanism causing the differences has been called rotation group bias. The causes of rotation group bias are not well understood, but the ways in which the survey conditions tend to change with the number of times in sample have been identified and include such effects as respondent and interviewer conditioning. Williams and Mallows (1970) have hypothesized that differential nonresponse by rotation group may cause the bias.

A second concern in the CPS is the increased use of telephone interviewing. It is known that differences in coverage, response rates, and responses to individual questions may occur between telephone and personal interviews. See, e.g., Groves (1977) and Bushery, et al., (1978). In the CPS, all first and fifth month interviews are to be conducted by personal visit. Ostensibly, all second month interviews are to be collected by personal visit, and all third, fourth, and sixth through eighth month interviews are to be collected by telephone when convenient. Telephone interviews, however, are used increasingly in all months because of rising costs of data collection.

A third possible problem is the respondent rule used in CPS. One person in the household, a responsible adult and generally the person who answers the door or the telephone, is chosen as respondent. There is no guarantee that the household respondent is the most knowledgeable member in the household regarding the employment status of all other individuals, and it may be that no one person in the household can satisfactorily

respond for all others. It may be that more accurate responses would be obtained if each person in the household responded for him or herself.

2.2 Experimental Treatments

In view of the nonsampling problems identified in Section 2.1, three experimental treatments were selected for study in the first phase of the MTP. They are 1) effect of continued interviewing by the same interviewer, 2) mode of interview, and 3) type of respondent. These treatments were selected because of their potential for understanding the cause of rotation group bias and the direct effect each of these may have on CPS estimates.

With respect to interview mode, the MTP is investigating the differences produced in the labor force data as a result of interviewing households by personal visit versus telephone. In any given month, half of the sample is designated to be interviewed by telephone, with the remaining half by personal visit. There are two exceptions to this procedure. First, all first month in sample households are enumerated by personal visit. In effect, this causes five-eights of the sample household to be assigned to the personal-visit mode with the remaining three-eights being assigned to the telephone mode. Second, in households assigned to be interviewed by a given mode, the other mode is allowed as a last resort to prevent loss of the interview.

Three levels of the type of respondent treatment are being used to investigate the accuracy of reporting by proxies and whether respondent conditioning to repeated interviews affects labor force responses. The first level uses the definition of household respondent, as currently practiced in CPS, in all interviews for the household. This treatment represents maximum conditioning of the respondent to the interview situation in that the respondent not only answers for him/herself each month, but also answers for all other eligible household members. The second alternative, involving the least conditioning, consists of the random designation of a household member each month to respond for him/herself and all other household members. Unless a household has only one eligible respondent, the designated respondent is not contacted in any two successive months. The third level is self-response, in which each eligible household member reports for him/herself each month. Certain deviations from the assigned type of respondent treatment are allowed in order to obtain an interview when otherwise one may be lost. A proxy interview is allowed in a self-response household when a respondent cannot be contacted or refuses to answer personally. Likewise, a self-response is allowed in a household assigned to be interviewed using the defined CPS household treatment or the designated respondent treatment when the respondent refuses to answer for another household member.

The interviewer assignment treatment tests possible conditioning effects by alternating interviewers in half the sample units, while the remaining half are enumerated by the same interviewer each month.

If interviewers or households are conditioned by repeated interviews, alternating the interviewers each month may result in a less marked pattern of rotation group bias.

Due to the rotation group structure of the MTP, time-in-sample may be regarded as a fourth experimental treatment. Four levels of this treatment (i.e., four rotation groups) are being used in the MTP, with a given group being enumerated for 4 consecutive months and then retired permanently. The scheme was chosen rather than the customary 4-8-4 pattern because of greatly reduced start-up time and because four rotation groups permit study of most of the rotation group bias effects found in CPS.

In addition to main effects, interactions between the four treatments are considered important in this study, especially between respondent and interviewer. Previous studies of CPS and other surveys have measured the effects of conditioning, differences in personal visit and telephone interviewing, and other factors, but little has been done to analyze interactions between different variables.

3. Sample and Experimental Design

3.1 Sample Design

The MTP experiments are being conducted in four primary sampling units (PSU's). They are the Los Angeles-Long Beach, California SMSA; the Chicago, Illinois SMSA; Lackawanna County, Pennsylvania; and Macon, Dooly, and Houston Counties, Georgia. These PSU's were selected purposively to display different types of unemployment problems, a mix of urban and rural characteristics, a representation of Blacks and Hispanics, and a wide geographic distribution. Another consideration that entered into the selection was the availability of field staff in the Census Bureau's 12 regional offices.

Within each PSU an unequal probability systematic sample of 32 1970 Census Enumeration Districts (ED's), or block groups, was chosen, with probability proportional to 1970 housing counts. All ED's, or block groups, within a given primary were sorted geographically prior to the systematic selection. In the Los Angeles and Chicago SMSA's two strata, consisting of the Central City and balance of SMSA, were used and the sample allocated equally to each. In the less urbanized areas, ED's were used, whereas block groups were selected in more urbanized areas. Each of the 32 ED's, or block groups, consists of approximately 250 housing units. The second stage units (SSU's) were then grouped into eight blocks of four using the order in which they were selected, each block being designated as a replicate in the design. Within each replicate, the second stage units were randomly assigned to one of four rotation groups.

The SSU's were canvassed before the initial enumeration in May 1978 with every housing unit in the area listed. Each month, one cluster of approximately 20 housing units is selected at random from each SSU. The listings for these clusters are updated the month before the clusters

are to come into sample to identify units which no longer exist (e.g., demolished). In the final stage, 12 units are subsampled from among the currently occupied units or those available for occupancy in each cluster and it is these units that are enumerated. The units in each rotation group remain in sample for 4 consecutive months; at that time a new sample of 12 units rotates in to take its place. In any 1 month, a total of 1,536 housing units are contacted.

The time of interviewing is the second and fourth weeks of each month. All units are randomly assigned to 1 of the 2 weeks in such a way that all treatments in the experimental design and all four rotation groups are equally represented in each week. In any 1 month, a total of 1,536 housing units are contacted.

Finally, it is important to note that the MTP sample is mutually exclusive of the CPS sample, and the MTP data are in no way used in preparing CPS estimates.

3.2 Experimental Design

The MTP may be viewed as a split-plot experiment. The whole plots are the second stage sampling units, i.e., ED or block group, and the whole-plot treatment is time-in-sample at four levels (i.e., first, second, third, or fourth month in sample). The split plots are the households. The split-plot treatments are the combinations of a complete $2 \times 2 \times 3$ factorial experiment, with interviewer-assignment at two levels, interview-mode at two levels, and respondent-type at three levels. One treatment combination is randomly assigned to each split plot in each whole plot.

In summary, there are 32 replicates in the entire MTP experiment with eight in each of four areas or PSU's. Each replicate contains four whole plots to which the levels of the treatment month-in-sample were randomly applied. There are 12 split plots in each whole plot, and 12 combinations of interviewer-assignment by interview-mode by type of respondent were assigned randomly to them. The total monthly sample of 1,536 housing units (32 replicates x 4 whole plots x 12 split plots) is accounted for in this way.

4. Proposed Analysis

There are a number of analyses planned for the MTP data, and two different analyses will be mentioned in this paper. The first is in the tradition of survey sampling, while the second is in the tradition of the general linear model. At this time we are only beginning the various analyses, and what is reported here is merely illustrative of the kinds of analyses that will eventually be made.

Before proceeding it is worth noting that no analysis of these data can measure the absolute size of estimator biases occurring from the use of the various experimental treatments. This would only be possible in the situation where an external source of validity was available, i.e., the true values of the various survey items were known. All that can legitimately be done in the MTP is measure differential biases and inter-actions between the treatments. While this is less than ideal, it does serve to broaden knowledge about survey errors and how alternative survey procedures interact with one another.

All analyses reported here are concerned with either the estimated unemployment rate or with the estimated proportion in the labor force. Other statistics are being analyzed or proposed for analysis, but they are not included in this article.

4.1 Analysis in the Tradition of Survey Sampling

The survey sampling approach is to make inference about treatment effects by looking at estimated contrasts, such as the difference $d = r_1 - r_2$, where r_1 is an estimated ratio (e.g., unemployment rate) based on observations from one treatment combination and r_2 is the comparable ratio estimated from another treatment combination. Ratios are estimated in the classical way, and reciprocals of the inclusion probabilities are used as weights in computing the numerator and denominator. In this approach, the estimator variance is estimated in accord with the sampling design and the form of the estimator. Then, Studentized statistics are used for testing hypotheses.

In our work to date, we have been computing jack-knife variance estimates. Eight pseudo-values are used, each being obtained by dropping one replicate from each of four primaries out of the sample. This way of computing the pseudo-values accounts for the multiple-stage design of the MTP. Then, $\hat{\sigma}_d^2$ denote the estimated variance of the estimator d, we have been making approximate tests using the Studentized statistic

$$t = d/\hat{\sigma}_d.$$

Table 1 presents some illustrative results. The first two columns give the two treatment combinations entering into the comparison; the third column gives the estimated difference d; and the fourth column gives the estimated standard error $\hat{\sigma}_d$. In making these computations, data were aggregated across the months June through November 1978, rather than making a separate analysis for each month. This procedure has the disadvantage of combining data from different time periods, and thus data which may have somewhat different moments. The main advantage of this analysis is that the sample sizes are effectively increased by aggregating across months, and this should result in more powerful tests.

As is evident from the table, most of the treatment differences are not significantly different from zero, and only two cases approach significance. First, the estimated t for the comparison (·, different, ·, designated) versus (·, same, ·, self) is -1.833. Second, the estimated t for the comparison (·, different, ·, designated) versus (·, same, ·, designated) is -1.846. By comparison with the percentage points of t_7 distribution, these differences are significant at about the $\alpha = .12$ level.

4.2 Analysis in the Tradition of the General Linear Model

As described in Section 3.2, the MTP was designed as a fully balanced split-plot experiment. The model is

$$Y_{ijk\ell m} = \mu + \rho_i + \mu_{jk\ell m} + u_{ijk\ell m},$$

where

$$\mu_{jk\ell m} = \alpha_j + \beta_k + \alpha\beta_{jk} + \gamma_\ell + \alpha\gamma_{j\ell} + \beta\gamma_{k\ell}$$
$$+ \alpha\beta\gamma_{jk\ell} + \delta_m + \alpha\delta_{jm} + \beta\delta_{km} + \alpha\beta\delta_{jkm}$$
$$+ \gamma\delta_{\ell m} + \alpha\gamma\delta_{j\ell m} + \beta\gamma\delta_{k\ell m} + \alpha\beta\gamma\delta_{jk\ell m}$$

and

$$u_{ijk\ell m} = \eta_{ij} + \varepsilon_{ijk\ell m}$$

for $i = 1, \ldots, 32$; $j = 1, \ldots, 4$; $k = 1, 2$; $\ell = 1, 2$; $m = 1, 2, 3$. In this notation, $Y_{ijk\ell m}$ is a response from the $ijk\ell m$-th household; μ is the overall mean; ρ_i is the effect of the i-th replicate; α_j is the effect of the j-th level of treatment A (i.e., month-in-sample); β_k is the effect of k-th level of treatment B (i.e., interviewer assignment); γ_ℓ is the effect of the ℓ-th level of treatment C (i.e., interview mode); δ_m is the effect of the m-th level of treatment D (i.e., type of respondent) and $\alpha\beta_{jk}$, $\alpha\gamma_{j\ell}$, $\beta\gamma_{k\ell}$, $\alpha\beta\gamma_{jk}$, $\alpha\delta_{jm}$, $\beta\delta_{km}$, $\alpha\beta\delta_{jkm}$, $\gamma\delta_{\ell m}$, $\alpha\gamma\delta_{j\ell m}$, $\beta\gamma\delta_{k\ell m}$, and $\alpha\beta\gamma\delta_{jk\ell m}$ are the various 2, 3, and 4 factor interactions. The classical fixed effects model is assumed. The error term has a one-fold nested structure where η_{ij} is the whole plot error and $\varepsilon_{ijk\ell m}$ is the split-plot error. As usual, it is assumed that the η_{ij} are independent $(0, \sigma_\eta^2)$ random variables, the $\varepsilon_{ijk\ell m}$ are independent $(0, \sigma_\varepsilon^2)$ random variables, and the η_{ij} and $\varepsilon_{ijk\ell m}$ are mutually independent.

Although the MTP was balanced by design, an unbalanced experiment is actually obtained because of missing observations (e.g., noninterviews) and certain interviewing procedures. For example, all first month in sample households are interviewed by personal visit, and the effect due to different interviewer assignments cannot be seen until the second month a household is in sample. This implies that any contrasts involving the following nine means are nonestimable: μ_{1121}, μ_{1122}, μ_{1123}, μ_{1211}, μ_{1212}, μ_{1213}, μ_{1221}, μ_{1222}, and μ_{1223}.

Because of the likelihood of significant interactions and the attendant difficulty in interpreting main effects, all analyses that we have made to date have used the $\mu_{jk\ell m}$ parameterization. In order to obtain a solution to the normal equations we have been imposing the nonestimable constraints

$$\mu = 0 \text{ and } \rho_1 = 0. \qquad (4.1)$$

Our analyses of these data have been performed using the software package SUPER CARP (cf.

Hidiroglou, Fuller, and Hickman (1978)). The computational algorithm is in two steps. First, the variance components are estimated based on an ordinary least squares fit. The model is then refit using generalized least squares on the estimated variance components.

We have been testing linear hypotheses of the general form H: $\underline{K}'\underline{\beta} = \underline{m}$ using the customary statistic $\hat{F} = (\underline{K}'\hat{\underline{\beta}} - \underline{m})' (\underline{K}'\hat{\underline{V}}\underline{K})^{-1} (\underline{K}'\hat{\underline{\beta}} - \underline{m})/s$ where $\underline{K}'$ is the matrix that defines the contrasts of interest, $\underline{m}$ is a vector of fixed constants, $\underline{\beta}$ is the vector of coefficients in the $\mu_{jk\ell m}$ parameterization, $\hat{\underline{\beta}}$ is the solution to the normal equations specified by (4.1), $\hat{\underline{V}}$ is the estimated covariance matrix of $\hat{\underline{\beta}}$ (i.e., using the generalized inverse corresponding to (4.1), and s is the row rank of $\underline{K}'$. Given the assumptions we have made, $\hat{F}$ is distributed approximately as an F with the usual degrees of freedom (cf. Fuller and Battese (1973)).

Some illustrative results are presented in Table 2. The dependent variable is proportion-in-civilian-labor-force for August 1978. $Y_{ijk\ell m}$ is the angular transformation of the proportion of individuals in the $ijk\ell m$-th household that are in the civilian labor force. The three tests presented here are for the main effect of month-in sample, and each is significant at the $\alpha = .10$ level. Evidently, the pattern is $\mu_{4\ldots} > \mu_{2\ldots} > \mu_{3\ldots} > \mu_{1\ldots}$. These tests should have a moderately clear interpretation because, although the results are not cited in this article, the three two-factor interactions involving month-in-sample are nonsignificant.

5. Discussion

This article is a preliminary report on the design, conduct, and analysis of the first phase of the Methods Test Panel (MTP). We discussed some of the potential nonsampling problems in the Current Population Survey (CPS), and showed how the MTP was designed to study the problems. At this time, we are only beginning to analyze these data. Some example results were included here merely to illustrate the kinds of analyses that we are starting to pursue.

In the future, we will be vigorously pursuing the analyses described in Section 4. This will be done for all months from June 1978 through November 1979, which is the last month of the first phase experiments, and will concentrate on several of the key labor force statistics such as the proportion in the civilian labor force, the unemployment rate, hours worked per employee, and so on. In addition to our analyses of the MTP responses, we are also planning analyses of the effects of the various treatments on the noninterview data. All of this work will be described in future reports.

Table 1. Selected Treatment Differences for the Unemployment Rate

Treatment combination 1[a]	Treatment combination 2[a]	$d=r_1-r_2$	$\hat{\sigma}_d$
(·, Different, ·, Household)	(·, Same, ·, Household)	.003	.008
(, Different, , Household)	(, Same, , Designated)	-.010	.011
(, Different, , Household)	(, Same, , Self)	-.008	.009
(, Different, , Designated)	(, Same, , Household)	-.011	.010
(, Different, , Designated)	(, Same, , Designated)	-.024	.013
(, Different, , Designated)	(, Same, , Self)	-.022	.012
(, Different, , Self)	(, Same, , Household)	.001	.011
(, Different, , Self)	(, Same, , Designated)	-.012	.014
(, Different, , Self)	(, Same, , Self)	-.010	.012
(, Different, , Self)	(, Different, , Household)	-.002	.012
(, Different, , Self)	(, Different, , Designated)	.012	.014
(, , , Designated)	(, , , Household)	-.001	.009
(, , , Self)	(, , , Household)	.004	.008
(, , , Self)	(, , , Designated)	.005	.010
(, , Personal,)	(, , Telephone,)	.006	.007
(, Same, , ,)	(, Different, ,)	.010	.008

[a]A treatment combination is of the form (level of month-in-sample, level of interviewer-assignment, level of interview-mode, level of respondent-type). The customary "dot" notation is used to denote averaging over all levels of a given treatment. In the case of month-in-sample, average is only over level 2, 3, and 4.

Source: Methods Test Panel data aggregated over months of June through November 1978.

Table 2. Tests for the Main Effect of Month-in-Sample for the Variable Proportion-in-Civilian-Labor-Force for August 1978

$$H: \quad \underset{\sim}{K}'\beta = (\mu_{1...}-\mu_{2...}) = 0$$
$$\underset{\sim}{K}'\hat{\beta} = -7.521$$
$$(\underset{\sim}{K}'\hat{\underset{\sim}{V}}\underset{\sim}{K})^{\frac{1}{2}} = 1.213$$
$$t = -6.200$$

$$H: \quad \underset{\sim}{K}'\beta = (\mu_{2...}-\mu_{3...}) = 0$$
$$\underset{\sim}{K}'\hat{\beta} = 1.450$$
$$(\underset{\sim}{K}'\hat{\underset{\sim}{V}}\underset{\sim}{K})^{\frac{1}{2}} = .748$$
$$t = 1.939$$

$$H: \quad \underset{\sim}{K}'\beta = (\mu_{3...}-\mu_{4...}) = 0$$
$$\underset{\sim}{K}'\hat{\beta} = -1.702$$
$$(\underset{\sim}{K}'\hat{\underset{\sim}{V}}\underset{\sim}{K})^{\frac{1}{2}} = .746$$
$$t = -2.282$$

NOTE: Computations are based on the customary angular transformation.

Bailar, Barbara A. (1975). The effects of rotation group bias on estimates from panel surveys. Journal of the American Statistical Association, 70, 23-30.

Bushery, John M., Cowan, Charles D. and Murphy, Linda R. (1978). Experiments in telephone-personal visit surveys. Proceedings Survey Research Methods Section, American Statistical Association.

Fuller, W.A. and Battese, G.E. (1973). Transformations for estimation of linear models with nested-error structure. Journal of the American Statistical Association 68, 626-632.

Gordon, R.A., et al., (1962). Measuring Employment and Unemployment. Report of the President's Committee to appraise employment and unemployment statistics. U.S. Government Printing Office, Washington, D.C.

Groves, Robert M. (1977). An experimental comparison of national telephone and personal interview surveys. Proceedings Social Statistics Section, American Statistical Association.

Hidiroglou, M.A., Fuller, W.A., and Hickman, R.D. (1978). SUPER CARP. Statistical Laboratory, Iowa State University, Ames, Iowa.

Waksberg, Joseph and Pearl, Robert B. (1965). New methodological research on labor force measuremtns. Proceedings Social Statistics Section, American Statistical Association.

Williams, W.H. and Mallows, C.L. (1970). Systematic biases in panel surveys due to differential non-response. Journal of the American Statistical Association 65, 1338-1349.