Survey Methodology, December 1996
Vol. 22, No. 2, pp. 147-155
Statistics Canada

The Application of McNemar Tests to the Current Population Survey's Split Panel Study

KATHERINE JENNY THOMPSON and ROBIN FISHER[1]

ABSTRACT

Results from the Current Population Survey split panel studies indicated a centralized computer-assisted telephone interviewing (CATI) effect on labor force estimates. One hypothesis is that the CATI interviewing increased the probability of respondent's changing their reported labor force status. The two sample McNemar test is appropriate for testing this type of hypothesis: the hypothesis of interest is that the marginal changes in each of two independent sample's tables are equal. We show two adaptations of this test to complex survey data, along with applications from the Current Population Survey's Parallel Survey split panel data and from the Current Population Survey's CATI Phase-in data.

KEY WORDS: Current Population Survey; Parallel survey; Nonparametric statistics.

1. INTRODUCTION

Results from the Current Population Survey's Parallel Survey split panel study and from the Current Population Survey's CATI Phase-in Project provided some indication of a centralized computer-assisted telephone interviewing (CATI) effect on the United States' monthly labor force estimates (Thompson 1994 and Shoemaker 1993). One hypothesis is that the CATI interviewing increased the probability of respondent's changing their reported labor force status from the first (personal) interview to the second (CATI) interview.

The two sample McNemar test is appropriate for testing this type of hypothesis. The McNemar test (1947) has been generalized to a two sample situation where the hypothesis of interest is that the marginal changes in each of two independent samples' 2×2 tables are equal (Feuer and Kessler 1989). The application presented was for a two sample cohort analysis and assumed simple random sampling.

Certain modifications of the test statistic for a McNemar test are necessary for a complex survey data application. First, because the data are not obtained through a simple random sample and are weighted, a separate estimate of the variance is required. Second, unless the survey has a longitudinal design, a separate link of individuals in two consecutive months' of data must be performed. In general, such a link will include some false matches and exclude some true matches. This adds another source of variance.

We show two adaptations of this test to complex survey data. In particular, we present these tests along with applications to the Current Population Survey's Parallel Survey split panel study and from the Current Population Survey's CATI Phase-in Project. In Section 2 we describe these test modifications including background on the one and two-sample McNemar tests (Section 2.1), modifications for complex survey data (Section 2.2), and some remarks on applications to several months' data (Section 2.3). Section 3 presents applications of these tests specifically to Current Population Survey Parallel Survey Data and to Current Population Survey CATI Phase-in data including background on the two studies (Section 3.1), details of the panel estimates and variance estimates (Section 3.2), diagnostics (Section 3.3), and results (Section 3.4). We make some concluding remarks in Section 4. Details of covariance estimation are included in the appendix.

2. TEST AND MODIFICATIONS

2.1 General

A sample is randomly split into two independent representative samples (split panels). After a baseline measurement is taken in both panels, a new technique is administered in one panel, the treatment panel. The other panel serves as a control.

The records are linked longitudinally after the second measured. A matched response can be +, −, or * (missing). Since this is matched data, the "**" cell will be empty.

This scenario is represented pictorially as

Treatment Panel
Month 2
Treatment

		+	−	*	
Month 1	+	x_{++}	x_{+-}	x_{+*}	$x_{+.}$
No Treatment	−	x_{-+}	x_{--}	x_{-*}	$x_{-.}$
	*	x_{*+}	x_{*-}		$x_{*.}$
		$x_{.+}$	$x_{.-}$	$x_{.*}$	n

[1] Katherine Jenny Thompson, Economic Statistical Methods and Programming Division, and Robin Fisher, Housing and Household Economic Statistics Division, United States Bureau of the Census, Washington, DC 20233, U.S.A.

Control Panel
Month 2
No Treatment

		+	−	*	
Month 1 No Treatment	+	x'_{++}	x'_{+-}	x'_{+*}	$x'_{+.}$
	−	x'_{-+}	x'_{--}	x'_{-*}	$x'_{-.}$
	*	x'_{*+}	x'_{*-}		$x'_{*.}$
		$x'_{.+}$	$x'_{.-}$	$x'_{.*}$	n'

where n is not necessarily equal to n'.

For each panel, define

$M_{(12)}$ as the set of cases which have month 1 and month 2 responses (matched cases). This set contains $n_{(12)} = (x_{++} + x_{+-} + x_{-+} + x_{--})$ elements;

$M_{(10)}$ as the set of cases which have month 1 responses, but no month 2 response. This set contains $n_{(10)} = (x_{+*} + x_{-*})$ elements;

$M_{(02)}$ as the set of cases which have month 2 responses, but no month 1 response. This set contains $n_{(02)} = (x_{*+} + x_{*-})$ elements.

Note that the n's are sample sizes and do not have weights.

First, consider the one-sample case. Traditionally, the one-sample McNemar test statistic is constructed from the $n_{(12)}$ and $n'_{(12)}$ matched responses, where a prime (′) indicates the control panel. In the one-sample scenario, we test the hypothesis

H_0: $p_{+-} = p_{-+}$, where the p's refer to cell probabilities
H_1: Not H_0

i.e., the hypothesis that the movement from one state to the other (+ to −, or − to +) is zero. We also refer to this movement as the flux.

The one-sample test can be a useful diagnostic in the two-sample situation. We examine the Control panel estimates to see if there is zero movement. Any significant movement in the Treatment panel can be measured as a deviation from zero flux or as a change in the probability of a "+."

The two-sample hypothesis is

H_0: $(p_{-+} - p_{+-}) = (p'_{-+} - p'_{+-})$
H_1: Not H_0.

In other words, the difference in the probabilities of switching in the two directions is the same, regardless of the treatment, or equivalently, the difference in panel fluxes is zero.

The Feuer and Kessler generalization (1989) to a two-sample McNemar test (described in 2.2.1 below) is confined to the $M_{(12)}$ and $M'_{(12)}$ linked sets. We can add an additional assumption, however, to allow the unmatched responses to be included in computation of the test statistics. This assumption motivates the discussion in Section 2.2.2.

2.2 Complex Survey Modifications

2.2.1 Modification One: Longitudinally Linked Data

This method is a straightforward application of the two-sample McNemar test, using longitudinally linked data from a complex survey.

To construct the test statistic, we examine the cell probabilities and note that

$$\begin{aligned}[p_{-+} - p_{+-}] &= [(p_{++} + p_{-+}) - (p_{++} + p_{+-})] \\ &= [p^*_{.+} - p^*_{+.}] \\ &= p^*_2 - p^*_1\end{aligned}$$

where p^*_2 is the marginal probability of a + response month 2, given a matched response for both months; and p^*_1 is the marginal probability of a + response month 1, given a matched response for both months.

The one-sample test statistic constructed from this panel's data is

$$Z^*_1 = \frac{p^\circ_2 - p^\circ_1}{\sqrt{\mathrm{Var}(p^\circ_2 - p^\circ_1)}},$$

where

$$p^\circ_1 = \frac{x_{++} + x_{+-}}{n_{(12)}}, \quad p^\circ_2 = \frac{x_{++} + x_{-+}}{n_{(12)}}.$$

Given two independent panels, the two-sample test statistic is

$$Z^* = \frac{(p^\circ_2 - p^\circ_1) - (p'^\circ_2 - p'^\circ_1)}{\sqrt{\mathrm{Var}(p^\circ_2 - p^\circ_1) + \mathrm{Var}(p'^\circ_2 - p'^\circ_1)}},$$

where

$$p'^\circ_1 = \frac{x'_{++} + x'_{+-}}{n'_{(12)}}, \quad p'^\circ_2 = \frac{x'_{++} + x'_{-+}}{n'_{(12)}}.$$

These results hold regardless of sample design. To extend the results to a complex survey application, we use weighted estimates and complex survey variances and covariances in place of simple random sample variances.

If the survey is designed to collect longitudinal data, then this modification is a natural extension of the method described by Feuer and Kessler. For this type of survey design, an effective mechanism to link individuals from month to month is presumably in place. Often, however, this is not the case, and one data set must be physically linked to another. Consequently, the $n_{(12)}$ elements in the domain will contain some false matches, and some actual matches may be inadvertently excluded. Both the record weights and variance estimates will need to be adjusted to account for the matching. Jabine and Scheuren (1986) provide an excellent summary of the methodological issues arising from the use of linked data, both for model-based and ad-hoc ("hard") record linkage techniques.

2.2.2 Modification Two: Unlinked Data

This method omits the longitudinal linkage step altogether, noting that the construction of the traditional McNemar test statistic can be expressed in terms of estimates of marginal probabilities. Assume that under the null hypothesis, the expected value of $(p_{*.} - p_{.*})$ is zero. This is described for a simple random sampling application in Marascuilo *et al.* (1988).

The domain for the first month of data is given $M_{(12)} \cup M_{(10)}$ which contains $n_{(12)} + n_{(10)} = n_1$ elements. The domain for the second month of data is given by $M_{(12)} \cup M_{(02)}$ which contains $n_{(12)} + n_{(02)} = n_2$ elements.

The one-sample test statistic constructed from the unlinked data is given by

$$Z_1 = \frac{p_2 - p_1}{\sqrt{\mathrm{Var}(p_2 - p_1)}},$$

where

$$p_1 = \frac{x_{+.}}{n_1}, \quad p_2 = \frac{x_{.+}}{n_2}.$$

Given two independent panels, the two-sample test statistic is

$$Z = \frac{(p_2 - p_1) - (p_2' - p_1')}{\sqrt{\mathrm{Var}(p_2 - p_1) + \mathrm{Var}(p_2' - p_1')}},$$

where

$$p_1' = \frac{x_{+.}'}{n_1'}, \quad p_2' = \frac{x_{.+}'}{n_2'}.$$

As with the application described in 2.2.1, all estimates are weighted estimates, and variances are complex survey variances.

2.3 Linear Combinations

We can use our estimated covariance matrix to test linear combinations of $\underline{\lambda}_T$, $\underline{\lambda}_C$, and $\underline{\delta}$ over time, where $\underline{\lambda}_T = \underline{p}_2 - \underline{p}_1$, $\underline{\lambda}_C = \underline{p}_2' - \underline{p}_1'$, and $\underline{\delta} = \underline{\lambda}_T - \underline{\lambda}_C$, and $\underline{p}_1, \underline{p}_2, \underline{p}_1'$ and $\underline{p}_2'$ are vectors containing the marginal probabilities for the time period under consideration.

General linear hypotheses of the form $\underline{K}'\underline{\mu}$ are now easily tested. One might wish to test for contrast by time period, for example testing the average difference from January through June against the remainder of the year's data. Perhaps the most interesting (to our applications) of these tests is of the hypothesis H_0: $\underline{1}'\underline{\mu} = 0$, where $\underline{\mu}$ is the expected value of one of the vectors described above.

Another test of particular interest is the "omnibus hypothesis," where we test H_0: $\underline{\mu} = \underline{0}$. The test statistics for this hypothesis are $\underline{\lambda}_T' \Sigma_{\lambda(T)}^{-1} \underline{\lambda}_T$, $\underline{\lambda}_C' \Sigma_{\lambda(C)}^{-1} \underline{\lambda}_C$, and $\underline{\lambda}_\delta' \Sigma_{\lambda(\delta)}^{-1} \underline{\lambda}_\delta$, each of which has an approximate chi-squared distribution with r degrees of freedom, where r is the dimension of the vector of interest.

3. APPLICATIONS

In this section, we apply the one and two-sample McNemar techniques for unlinked data outlined in 2.2.2 and 2.3 to two separate sets of data: the Current Population Survey's Parallel Survey split panel data and Current Population Survey CATI Phase-in data. Tables 1 and 2 (section 3.4.1) provide the results for Parallel Survey split panel data. Tables 3 and 4 (section 3.4.2) provide the results for the Current Population Survey CATI Phase-in data.

3.1 Background

The official monthly civilian labor force estimates from January 1994 onward are based on data from a comprehensively redesigned Current Population Survey. The redesign included implementation of a new, fully computerized questionnaire, and an increase in centralized computer-assisted telephone interviewing (CATI). To gauge the effect of the Current Population Survey redesign on published estimates, a Parallel Survey was conducted using the new questionnaire and data collection procedures from July 1992 through December 1993. Special studies were embedded in both the Parallel Survey and the Current Population Survey during the same time period to provide data for testing hypotheses about the effects of the new methodological differences on labor force estimates: the Parallel Survey split panel study and the Current Population Survey CATI Phase-in Project (a continuation of the study presented in Shoemaker 1993).

The effect of increased centralized computer-assisted telephone interviewing was of particular interest. Findings from the study described in Shoemaker (1993) had shown that including centralized telephone interviews tended to yield a larger unemployment rate. The two-sample McNemar test appeared to be a good vehicle for examining this phenomenon. In both the Current Population Survey and the Parallel Survey, households are interviewed for 4 consecutive months, not interviewed for the next 8 consecutive months, and then interviewed for another 4 consecutive months. The first and fifth interviews are conducted by a personal visit, and the subsequent interviews are conducted by telephone whenever possible. Thus the first and fifth interviews provide a baseline measurement of labor force status; the second and sixth interviews provide a "post-treatment" measurement of labor force status.

To create the panels for both studies, sample within selected sample areas was randomly divided into two representative panels using systematic sampling methods. The treatment panel was designated as CATI eligible. This meant that the sample households in the panel were eligible for interview at a centralized facility after the initial (first and fifth) interviews. To be interviewed by CATI, a respondent must have a telephone and speak English or Spanish, and must agree to be interviewed in subsequent months by telephone. Not all households in this panel were interviewed by CATI. The other panel served as a control.

The monthly unemployment rate is the primary statistic of interest published from Current Population Survey data. This rate is defined as the estimated number of unemployed persons divided by the estimated number of persons in the civilian labor force (the denominator does not include military personnel, persons under sixteen years old, or people who are no longer looking for work, or retired persons). Our primary goal was to understand how including CATI interviews influenced the probability of changing labor force status, in this case from unemployed to not unemployed (or vice versa). Our statistics for the one and two-sample McNemar tests used unemployment to population ratios, rather than unemployment rates. This allowed for a slightly more precise estimate of the proportion by decreasing the variability of the test statistic.

3.2 Estimates

Each month/panel estimate is an unbiased estimate. That is, the weights used to produce the estimates were strictly a function of the probability of selection: each weight is the product of the baseweight (the inverse probability of selection for a PSU), the weighting control factor (an adjustment for field subsampling), and a split panel factor (an adjustement for the probability of inclusion in a split panel). The split panel factor for the Parallel Survey study was constant by design: nine tenths of the sample was randomly assigned to the treatment panel. The split panel factors for the CPS CATI Phase-in were not constant: the sample in the treatment panel varied on a monthly level, as more sample was randomly assigned to CATI facilities.

Variances of levels were computed with generalized variance functions (GVFs). For more details, see Fisher *et al.* (1993). Robert Fay used his VPLX software (Fay 1990) to calculate replicate estimates of correlation between rotation groups for unemployed and for civilian labor force using September 1992 through December 1993 data from the Current Population Survey. We used these correlations for the test statistics based on unlinked data, assuming that they would not differ by survey (Current Population Survey versus Parallel Survey) or by geography (national versus subnational). We derived an expression for the within-panel correlation for civilian population by relating previously calculated autocorrelations (Fisher and McGuinness 1993) and variance estimates to the individual rotation group estimates. See the appendix for details of the estimation of the correlations.

We did not use the linked modification in our applications for several reasons. The primary reason was the difficulty of longitundinally matching the data. Moreover, we were unable to evaluate the success of our matching. Finally, we did not have any estimates of correlation for the linked data.

Implicit in our analysis of the unlinked data is the assumption that the probability of a nonresponse (or a nonmatch) is random. We assume that the probability of a nonresponse one month is independent of the respondent's labor force classification in the previous month. This assumption is not universally recognized. In fact, Stasny and Fienberg (1984) argue the reverse, and propose several alternative discrete-time models for the use of longitudinally linked CPS data. In our application, the estimates of marginal probabilities based on our (perhaps) poorly matched linked data were almost identical to the estimates based on unlinked data, and so we feel that our analysis did not suffer particularly from our assumption.

3.3 Diagnostics

Small expected sample sizes in individual cells will result in highly variable and consequently unreliable tests. We are not aware of a general method of calculating adequate sample sizes for this type of analysis using complex survey data. Instead, as a naive approach we used a slightly modified version of the traditional Pearson chi-squared test diagnostic to form a cut-off value as follows:

As defined in Section 2.2.2, let

$x_{+.}$ = unweighted unemployed persons in month 1;
$x_{-.}$ = unweighted not-unemployed persons in month 1;
$x_{.+}$ = unweighted unemployed persons in month 2;
$x_{.-}$ = unweighted not-unemployed persons in month 2.

Recall that in the case of the usual contingency table, $E[+-] = x_{+.}x_{.-} / n_{(12)}$, $E[-+] = x_{-.}x_{.+} / n_{(12)}$ under the assumption of independence (and ignoring missing values). In our estimates of expected cell size, we used unlinked marginal data. The sample sizes for the two margins corresponding to the two months are different; that is, the denominators of the expected cell totals are different depending on which margin we examine. Because we could not observe $n_{(12)}$, we estimated it by the geometric mean of n_1 and n_2, which seemed to most closely resemble the expression for the expected cell size. We have not evaluated the effectiveness of the geometric mean versus alternative estimators.

A commonly used rule in contingency table analysis is that expected cell sizes should be at least five. However, both the Current Population Survey and Parallel Survey designs are highly clustered, and we felt that the cut-off value should be adjusted upwards. Accordingly, we multiplied the cut-off value by a design effect. We further increased the cut-off value for expected cell sizes to compensate for the correlation between the rows and columns of our tables to arrive at our final cut-off expected cell size of ten.

3.4 Results

3.4.1 Parallel Survey Split Panel Study

This section presents the formal results from the one and two-sample McNemar tests using unlinked Parallel Survey split panel data. Although this data was collected monthly, small expected cell sizes in the control panel led us to omit several sets of adjacent months from this analysis. Table 1

Table 1
One-Sample McNemar Tests for Individual Parallel Survey Panels – Unlinked Data

Time Frame	Treatment Panel			
	$p_2 - p_1$	se$(p_2 - p_1)$	Z-Statistic	P-Value
10/92 – 11/92	-0.62	0.29	-2.18	0.03
11/92 – 12/92	-0.47	0.28	-1.68	0.09
04/93 – 05/93	-0.76	0.27	-2.84	0.00
06/93 – 07/93	-0.04	0.27	-0.16	0.88
08/93 – 09/93	-0.66	0.27	-2.42	0.02
	Control Panel			
	$p_2' - p_1'$	se$(p_2' - p_1')$	Z-Statistic	P-Value
10/92 – 11/92	2.44	0.81	3.02	0.00
11/92 – 12/92	0.11	0.83	0.14	0.89
04/93 – 05/93	0.20	0.72	0.27	0.78
06/93 – 07/93	0.97	0.71	1.38	0.17
08/93 – 09/93	-1.73	0.68	-2.54	0.01

provides summary statistics for the one-sample "monthly" tests for each panel which were based on unlinked data from the Parallel Survey's split panels. Table 2 provides summary statistics for the two-sample tests based on unlinked data.

The reported values of p_1, p_2, p_1', and p_2' are percentages of estimated unemployed to estimated total population for the panel. Recall that p_1 and p_1' are the panel ratio of estimated unemployed from the first and fifth interviews to the estimated panel population from the first and fifth interviews; p_2 and p_2' are the panel ratio of estimated unemployed from the second and sixth interviews to the estimated panel population from the second and sixth interviews. Data from the time frame of February 1993 – March 1993 are omitted: a CATI facility was closed during the March interview week because of a blizzard.

The one-sample McNemar tests in Table 1 test the probability that the proportion unemployed does not change between the initial and the subsequent interview within the same panel. We use the Control panel to examine the unemployment flux from one month to the next in the absence of CATI. Note that the two significant point estimates are in the opposite direction.

The entire vector of differences of proportions was found to be significantly different from the zero vector (p-value = 0.00), but the sum of the individual components was not found to be significant (p-value = 0.24). Consequently, we did not test any further linear combinations.

We expected a certain amount of month-in-sample bias to be present in these estimates. In Adams (Bureau of the Census 1991), the estimates of p_1 constructed from the first and fifth months in sample of the full Current Population Survey were roughly six percent larger than their respective second and sixth month-in-sample analogues (p_2). Consequently, estimates of $(p_2 - p_1)$ calculated from the full Current Population Survey data were generally negative. As seen in Table 1, this was not the case with the Parallel Survey Control panel's estimates: counter to our intuition, the estimated difference $(p_2' - p_1')$ is generally positive. This could be a function of the time difference, a geographic difference, or a design difference. Adams used 1987 data from the Current Population Survey to calculate national estimates of biases associated with rotation groups. Thus in each of these one-sample tests, the net movements are intertwined with an unmeasured effect from month-in-sample bias.

Note the negative unemployment flux in the Treatment panel. This observation is supported by the significant result from the formal test of the omnibus hypothesis (p-value = 0.00), and the significant result for the hypothesis $\underline{1}'\underline{\mu} = 0$ (p-value = 0.00).

The two-sample McNemar test results are presented below.

Table 2
Two-Sample McNemar Tests – Unlinked Parallel Survey Data

Time Frame	$(p_2 - p_1) - (p_2' - p_1')$	se$[(p_2 - p_1) - (p_2' - p_1')]$	Z-Statistic	P-Value
10/92 – 11/92	-3.06	0.86	-3.58	0.00
11/92 – 12/92	-0.58	0.88	-0.66	0.51
04/93 – 05/93	-0.95	0.77	-1.24	0.22
06/93 – 07/93	-1.02	0.76	-1.34	0.18
08/93 – 09/93	1.08	0.74	1.47	0.14

Individually, the monthly results do not demonstrate a clear difference in the unemployment flux between the two panels. On the other hand, the omnibus test statistic is significant (p-value = 0.00). The mean unemployment flux seems to be lower in the treatment panel as evidenced by the significant test results of the hypothesis $\underline{1}'\underline{\mu} = 0$, where $\underline{\mu}$ is the vector of $((p_2 - p_1) - (p_2' - p_1'))_i$'s, with each element corresponding to a month's estimate (p-value = 0.01).

In these tests, we make statements about contrasts in a table of probabilities, looking for indicators of the effect of a treatment on unemployment movement. As mentioned earlier, some month-in-sample bias is present in the one-sample tests. The tested hypotheses examine combinations of the net movement within a panel and month-in-sample bias. This problem is somewhat mitigated in the two-sample tests. Indeed, if month-in-sample bias is an additive term which affects both panels equally, it will cancel out of the test statistic. Moreover, this effect will be alleviated somewhat in the two-sample test even if it is not the same between the two panels or is multiplicative. Our preliminary sensitivity analysis bore this out: we found that the one-sample tests were sensitive to month-in-sample bias, but that the two-sample tests were not.

The two-sample t-tests presented in Thompson (1994) failed to detect a difference by panel in mean unemployment rate using the Parallel Survey split panel data. This contrasts with the Current Population Survey CATI Phase-in results: over two years, the CATI (Treatment) panel had consistently significantly higher unemployment rates than the non-CATI

(Control) panel. See Shoemaker (1993). In this analysis of Parallel Survey split panel data, we have evidence that the expected value of the proportion unemployed is lower in the presence of CATI. There are, however, some problems with the data. First, as previously mentioned, there is some confounding in the Treatment (CATI) panel, since not all respondents in this panel have their second interview conducted from a centralized telephone facility. Second, in each month the expected sample size in the Control panel cells was near ten, which could be small enough to make the distribution behave unpredictably. This latter problem is not an issue with the Current Population Survey CATI Phase-in study analysis presented in 3.4.2.

3.4.2 Current Population Survey CATI Phase-in Project Results

The Current Population Survey CATI Phase-in project was a continuation of the study presented in Shoemaker (1993). The primary purpose of this study was to measure the effect of including CATI interviewing on the unemployment rate. CATI interviewers in this study used an automated version of the old Current Population Survey paper questionnaire, which had a slightly modified version of the lead-in labor force question. More details are provided in Thompson (1994). The data considered in this paper are from the same time period as the Parallel Survey split panel data examined in 3.4.1: October 1992 through December 1993, again omitting the February 1993 – March 1993 time frame. Expected cell sizes in both the Treatment (CATI) and Control (non-CATI) panels were well over one hundred, and so all other contiguous months of data are included.

The one-sample McNemar test results for both panels are presented in Table 3. Test statistics are constructed with unlinked data. The reported values of p_1, p_2, p_1', and p_2' are percentages of estimated unemployed to estimated total population for the panel.

Table 3
One-Sample McNemar Tests for Individual Current Population Survey Panels – Unlinked Data

Time Frame	Treatment Panel			
	$p_2 - p_1$	$se(p_2 - p_1)$	Z-Statistic	P-Value
10/92 – 11/92	1.13	0.16	7.63	0.00
11/92 – 12/92	0.07	0.17	0.44	0.66
12/92 – 01/93	0.43	0.13	3.46	0.00
01/93 – 02/93	0.00	0.14	0.03	0.97
03/93 – 04/93	-0.25	0.14	-1.81	0.07
04/93 – 05/93	0.63	0.13	4.99	0.00
05/93 – 06/93	0.88	0.13	6.56	0.00
06/93 – 07/93	0.84	0.13	6.49	0.00
07/93 – 08/93	-0.07	0.14	-0.51	0.61
08/93 – 09/93	0.42	0.13	3.17	0.00
09/93 – 10/93	0.06	0.12	0.52	0.60
10/93 – 11/93	1.05	0.12	8.45	0.00
11/93 – 12/93	0.18	0.14	1.27	0.20
	Control Panel			
	$p_2' - p_1'$	$se(p_2' - p_1')$	Z-Statistic	P-Value
10/92 – 11/92	0.05	0.47	0.11	0.92
11/92 – 12/92	-0.14	0.47	-0.30	0.76
12/92 – 01/93	0.72	0.43	1.68	0.09
01/93 – 02/93	-0.91	0.43	-2.11	0.03
03/93 – 04/93	-0.16	0.39	-0.40	0.69
04/93 – 05/93	-0.18	0.43	-0.42	0.67
05/93 – 06/93	0.47	0.38	1.22	0.22
06/93 – 07/93	-0.32	0.46	-0.68	0.49
07/93 – 08/93	-0.52	0.39	-1.32	0.19
08/93 – 09/93	-0.54	0.44	-1.21	0.23
09/93 – 10/93	-0.08	0.37	-0.22	0.83
10/93 – 11/93	-0.63	0.42	-1.50	0.13
11/93 – 12/93	-0.09	0.37	-0.23	0.82

As with the Parallel Survey split panel data, the one-sample McNemar tests using the CATI Phase-in data test the probability that the proportion unemployed does not change between the initial and the subsequent interview within the same panel. Again, we use the Control panel to estimate the unemployment flux from one month to the next in the absence of CATI. The monthly tests for the Control panel do not appear to exhibit any particular movement. Furthermore, the omnibus hypothesis test was not significant (p-value = 0.29), so we did not test any further linear combinations.

Again basing our expectations on the effects of month-in-sample bias presented in Adams (1991), we believed that the Control panel estimate of p_1' (from the first and fifth months-in-sample) would be larger than its respective second and sixth month-in-sample analog, p_2'. On the average, this was the case: although quite variable, the estimates of p_1' are on the average about 4 percent larger than the estimates of p_2'. Because both panels are representative samples from the same parent sample, we assume that the month-in-sample bias behaves similarly in both panels. The Treatment (CATI) panel estimates of p_2 are *larger* on the average than the estimates of p_1. Given the Control panel's estimates behavior, this phenomenon provides some evidence of a CATI effect.

Note the movement in the Treatment panel from *not* unemployed to unemployed. This observation is supported by the significant result from the formal test of the omnibus hypothesis (p-value = 0.00), and the significant result for the hypothesis $\underline{1}'\underline{\mu} = 0$ (p-value = 0.00). In contrast to the Parallel Survey results provided in 3.4.1, this data provides some evidence that unemployment rate is higher in the presence of CATI. This evidence is further supported by the two sample McNemar test results provided Table 4. The individual monthly results in Table 4 provide some evidence of difference in the unemployment flux between two panels. Furthermore, the omnibus test is significant (p-value = 0.00). The mean unemployment flux in the Treatment panel seems to be higher as evidenced by the significant test results of the hypothesis $\underline{1}'\underline{\mu} = 0$.

The two-sample t-tests presented in Thompson (1994) also detected a *positive* difference by panel in mean unemployment rate using the Current Population Survey split panel data

Table 4
Two-Sample McNemar Tests – Unlinked Current Population Survey Data

Time Frame	$(p_2 - p_1) - (p_2' - p_1')$	$se[(p_2 - p_1) - (p_2' - p_1')]$	Z-Statistic	P-Value
10/92 – 11/92	1.18	0.50	2.38	0.02
11/92 – 12/92	0.22	0.50	0.43	0.67
12/92 – 01/93	-0.29	0.45	-0.64	0.52
01/93 – 02/93	0.92	0.45	2.03	0.04
03/93 – 04/93	-0.10	0.42	-0.23	0.81
04/93 – 05/93	0.81	0.45	1.81	0.07
05/93 – 06/93	0.41	0.41	1.01	0.31
06/93 – 07/93	1.16	0.48	2.41	0.02
07/93 – 08/93	0.45	0.42	1.07	0.28
08/93 – 09/93	0.95	0.46	2.06	0.04
09/93 – 10/93	0.14	0.39	0.37	0.71
10/93 – 11/93	1.69	0.44	3.83	0.00
11/93 – 12/93	0.26	0.40	0.66	0.51

i.e., including CATI interviews resulted in a *higher* unemployment rate. These results were consistent with the Current Population Survey CATI Phase-in results presented in Shoemaker (1993). This analysis of Current Population Survey split panel data reinforces that conclusion. Again, it is impossible to attribute the positive net migration from not unemployed to unemployed entirely to the effect of CATI: the same confounding described in 3.4.1 is present in this Treatment (CATI) panel.

3.5 Discussion

Our results appear to yield opposite conclusions about the effect of CATI on unemployment flux. The CATI effect is not, however, the same in both tests.

Perhaps the key difference is the questionnaire. The Parallel Survey data was collected using the newly redesigned Current Population Survey questionnaire. The new questionnaire was designed as an automated instrument. In contrast, the old Current Population Survey questionnaire used for the Current Population Survey CATI Phase-in Project was designed as a paper instrument. Field interviewers were required to memorize complicated skip patterns. To minimize respondent burden, both versions of the Current Population Survey questionnaire are designed for an average interview length of twenty minutes. Using an automated questionnaire, an interviewer can collect more (and more detailed) information in the same amount of time, since she no longer has to determine the path of the interview. Besides the automation difference, the wording of the labor force questions differs between the two questionnaires.

Parallel Survey interviews were conducted using the same questionnaire both in the field interviews (using a laptop computer) and in the CATI facilities. In contrast, the Current Population Survey CATI Phase-in interviews used two different versions of the old questionnaire: a paper version for the field interviews; and an automated version, with a slightly modified lead-in labor force question for the CATI interviews.

Given these questionnaire differences, and the caveats about the Parallel Survey split panel data, we view our results as preliminary. Instead, we recommend pursuing this examination using one and two-sample McNemar techniques on the new Current Population Survey split panel data, which uses the old CATI Phase-in design and the redesigned, fully automated questionnaire.

4. CONCLUSION

We have presented two modifications of the one and two-sample McNemar tests using complex survey data, with applications from the unlinked data modification. If the survey does not have a longitudinal design, then the application using the linked data will have an unknown variance/covariance structure and will include a variance component due to matching error. In this case, using the unlinked data makes sense with respect to the model's interpretation, although the statistic based on the (unlinked) estimates of marginal probabilities may be inferior to a well-developed linked model. If the survey has a longitudinal design, then the first method may be preferred, as it is a straight-forward extension of the traditional test, and consequently, the interpretation is equivalent to the textbook interpretation.

The two-sample McNemar test is not the sole approach one might use in the situation described in section 2.2.2. Another approach to the unlinked form of this problem would be to use a log-linear model for a $2 \times 2 \times 2$ contingency table as in Rao and Scott (1984). In either case, there are trade-offs. The interpretation of the McNemar test is intuitive: it is a cause and effect model, or a repeated measures type of experimental design. The $2 \times 2 \times 2$ contingency table model's interpretation is perhaps less intuitive. Note, however, that the test statistic for the McNemar tests are "Wald-like" statistics, which are often considered to be less efficient than the chi-squared type, *e.g.*, Fay (1985). It is also worth noting that unlike the Rao-Scott formulation, the approach described in this paper makes explicit provisions for the use of linked data.

Areas for future research include investigations into the power of these tests in the context of complex sample data, variance/covariance estimation for linked data including matching error variance contributions, and the difference in efficiency in the two approaches. In data analytical applications, one and two-sample McNemar tests seem to have uses in comparing aspects of different survey methods or effects on responses within a method over time. The approach is nonparametric in its conception; when the approximation is good, it avoids pitfalls that may be associated with model-based tests.

ACKNOWLEDGEMENTS

We thank James Hartman, Alfredo Navarro, James Roebuck, Lynn Weidman, the referee, and the editors for their useful comments. We also thank Sue Chandler for her typing

of this paper. This paper reports the general results of research undertaken by Census Bureau staff. The views expressed are attributable to the authors and do not necessarily reflect those of the Census Bureau.

APPENDIX

For the unlinked data modification of the McNemar Test, $(p_2 - p_1)$ is estimated by $X_{+.}/N_1 + X_{.+}/N_2$ where $X_{+.}$, $X_{.+}$, N_1, and N_2 are weighted estimates, and

$$\widehat{\mathrm{Var}}(p_2 - p_1) = \left[\frac{X_{+.}}{N_1}\right]^2 \left[\frac{\mathrm{Var}(X_{+.})}{X_{+.}^2} - \frac{\mathrm{Var}(N_1)}{N_1^2}\right] + \left[\frac{X_{.+}}{N_2}\right]^2 \left[\frac{\mathrm{Var}(X_{.+})}{X_{.+}^2} - \frac{\mathrm{Var}(N_2)}{N_2^2}\right] - \frac{2X_{+.}X_{.+}}{N_1 N_2}\left[\frac{\mathrm{Cov}(X_{+.}X_{.+})}{X_{+.}X_{.+}} - \frac{\mathrm{Var}(N_1)}{N_1^2} - \frac{\mathrm{Var}(N_2)}{N_2^2} + \frac{\mathrm{se}(N_1)\,\mathrm{se}(N_2)}{N_1 N_2}\right].$$

In this appendix we discuss the derivation of the covariance term in the variance estimate, considering only the unlinked data.

Consider the within-panel correlation

$$\mathrm{Cov}(X_{+.}, X_{.+}) = \mathrm{Cov}\left(\sum_{j=1,5} X_{1,j}, \sum_{j=2,6} X_{2,j}\right) \tag{A1}$$

where $X_{i,j}$ is a weighted sample level for month i, month-in-sample (MIS) j. Note that $X_{1,j}$ and $X_{2,j+1}$ are from the same rotation group unless $j = 4$ since a rotation group is out of sample for eight months after being in for four.

We assumed that the correlations between $X_{i,j}$ and $X_{k,m}$ can be decomposed into three separate categories:

1) A within-rotation-group correlation,

$$\mathrm{Cov}(X_{i,j}, X_{i+1,j+1}) = r_1, \text{ when } j = 1,2,3,5,6,7.$$

2) A within-month-between-rotation group correlation,

$$\mathrm{Cov}(X_{i,j}, X_{i,k}) = \omega, \ k \neq j, \text{ and}$$

3) A between-rotation-group between-month correlation.

$$\mathrm{Cov}(X_{i,j}, X_{i+1,k}) = \gamma, \ k \neq j+1 \text{ or } j = 3.$$

Replicate estimates of these correlations were available.

The covariance in (A1) becomes

$$\begin{aligned}\mathrm{Cov}(X_{+.}, X_{.+}) &= \mathrm{Cov}(X_{1,1} + X_{1,5}, X_{2,2} + X_{2,6}) \\ &= \mathrm{Cov}(X_{1,1}, X_{2,2}) + \mathrm{Cov}(X_{1,1}, X_{2,6}) + \\ &\quad \mathrm{Cov}(X_{1,5}, X_{2,2}) + \mathrm{Cov}(X_{1,5}, X_{2,6}) \\ &= 2(r_1 + \gamma)\,\mathrm{Var}(X_{i,j}),\end{aligned} \tag{A2}$$

using the simplifying assumption that $\mathrm{Var}(X_{i,j})$ is constant for all i and j. The variance for a full month's estimate, $\mathrm{Var}(\sum_{j=1}^{8} X_{i,j})$ is available in the form of a generalized variance function (GVF). We use this estimate to calculate $\mathrm{Var}(X_{i,j})$ by applying the following derivation:

$$\begin{aligned}\mathrm{Var}\left(\sum_{j=1}^{8} X_{i,j}\right) &= \sum_j \sum_k \mathrm{Cov}(X_{i,j}, X_{i,k}) \\ &= \sum_j \mathrm{Var}(X_{i,j}) + \sum_{j \neq k} \mathrm{Cov}(X_{i,j}, X_{i,k}) \\ &= (8 + 56\omega)\,\mathrm{Var}(X_{i,j})\end{aligned}$$

so

$$\mathrm{Var}(X_{i,j}) = (8 + 56\omega)^{-1}\,\mathrm{Var}\left(\sum_{j=1}^{8} X_{i,j}\right). \tag{A3}$$

REFERENCES

ADAMS, D.E. (1991). Current population survey month-in-sample (MIS) bias index research. Internal memorandum, Demographic Statistical Methods Division, U.S. Bureau of the Census, Washington, DC.

DONNER, A., and LI, K.Y.R. (1990). The relationship between chi-square statistics from matched and unmatched analyses. *Journal of Clinical Epidemiology*, 43, 827-831.

FAY, R. (1990). VPLX: Variance estimates for complex samples. *Proceedings of the Section on Survey Research Methods, American Statistical Association.*

FAY, R. (1985). A jackknifed chi-squared test for complex samples. *Journal of the American Statistical Association,* 80, 148-157.

FEUER, E.J., and KESSLER, L.G. (1989). Test statistic and sample size for a two-sample McNemar test. *Biometrics*, 45, 629-636.

FISHER, R., ROBISON, E., THOMPSON, J., and WELCH, M. (1993). Variance estimation in the current population survey overlap test. *Proceedings of the Section on Survey Research Methods, American Statistical Association.*

FISHER, R., and McGUINNESS, R. (1993). Correlations and adjustment factors for current population survey. Internal memorandum, Demographic Statistical Methods Division, U.S. Bureau of the Census, Washington, DC.

HOGUE, C. (1984). History of the problems encountered estimating gross flows. *Proceedings of the Conference on Gross Flows in Labor Force Statistics,* 1-7.

JABINE, T.B., and SCHEUREN, F.J. (1986). Record linkages for statistical purposes: Methodological Issues. *Journal of Official Statistics*, 2, 3, 255-277.

McNEMAR, Q. (1947). Note on the sampling error of the differences between correlated proportions of percentages. *Psychometrika*, 12, 153-157.

MARASCUILO, L.A., OMELICH, C.L., and GOKHALE, D.V. (1988). Planned and post hoc methods for multiple-sample McNemar (1947) tests with missing data. *Psychological Bulletin*, 103, 238-245.

RAO, J.N.K., and WU, C.F.J. (1988). Resampling inference with complex survey data. *Journal of the American Statistical Association*, 83, 231-241.

RAO, J.N.K., and SCOTT, A.J. (1984). On chi-squared tests for multiway contingency tables with cell proportions estimated from survey data. *The Annals of Statistics,* 12, 46-60.

SHOEMAKER, H.H. (1993). Results from the current population survey CATI phase-in project. *Proceedings of the Section on Survey Research Methods, American Statistical Association.*

STASNY, E.A., and FIENBERG, S.E. (1984). Some stochastic models for estimating gross flows in the presence of nonrandom nonresponse. *Proceedings of the Conference on Gross Flows in Labor Force Statistics*, 25-39.

THOMPSON, J. (1994). Mode effects analysis of labor force estimates. Current Population Survey Overlap Analysis Team Technical Report 3, U.S. Bureau of the Census, Washington, DC.

THOMPSON, J., and FISHER, R. (1994). Two sample McNemar test for complex surveys. *Proceedings of the Section on Survey Research Methods, American Statistical Association.*