MEASURING RESPONSE ERRORS IN SURVEY DATA

Michael D. Sinclair[1], U.S. Bureau of the Census, and
Joseph L. Gastwirth[2], George Washington University
Michael Sinclair, U.S. Bureau of the Census, Washington D.C. 20233

Key Words: Reinterview Surveys, Response Bias

1. Introduction

Reinterview surveys are often conducted to measure the level of content errors in the data resulting from a survey or a census. Content errors, often referred to as response errors, are reported deviations from the respondent's true values. The expectation of the distribution of the these random response errors is the level of response bias associated with the data. The variability of the distribution is defined as the response variance. These two parameters represent the level, respectively of both systematic and variable errors in the recorded data and are used extensively as an assessment of data quality.

Two different reinterview designs are generally implemented independently to measure response bias and variance as discussed by U.S. Census Bureau (1985). In the response-bias type study, the reinterview survey is designed to be more accurate through the use of "preferred" data collection procedures than the parent survey methods. Preferred procedures often include the use of highly qualified or more experienced staff or extensive probing techniques. In a response-variance type study, the reinterview attempts to replicate the essential parent survey conditions.

Hui and Walter (1980) have demonstrated a methodology for evaluating the accuracy of a new diagnostic test against a standard test with unknown error rates when both tests were applied to a sample of patients. By treating the reinterview as the standard test and the parent survey as the new procedure, these methods can be applied to evaluate the level of classification error in both the parent survey and the reinterview. These procedures are limited in their applicability to certain demographic or economic groups, requiring the analyst to be able to divide the data into two subpopulations which have different prevalence rates for a characteristic, but equivalent misclassification errors from the parent-reinterview survey process. An example of the application of the Hui and Walter methods to the Census Bureau's Content Reinterview Survey (CRS) program will be presented.

2. Matched Parent-Reinterview Survey Data Models

In general, consider a sequential application of 2 tests to a sample of n sample units. Denote each testing instrument by the value of r, with $r = 1$ corresponding to the parent survey responses for the sample units in the reinterview and $r = 2$ for the reinterview responses. Each test classifies each sample unit into one of two possible response categories (or the classification is collapsed into two categories of interest) resulting in a 2^2 contingency table for the test outcomes. The true classification status of each sample unit is unknown and each of the tests has unknown levels of classification error.

Figure 1

Parent Survey Responses	Reinterview Survey Responses		
	Positive	Negative	Total
Positive	A	C	A+C
Negative	B	D	B+D
Total	A + B	C + D	n

In the table presented in Figure 1 the letters A, B, C, D represent the multinomial frequencies for each of the possible outcomes. Assume that the frequencies are limited to those that belong to an arbitrary subpopulation g of the sample units. Define π_g as the true prevalence rate of the sample units in subpopulation g and let $\alpha_{r,g}$ and $\beta_{r,g}$ denote the false positive and false negative rates, respectively, associated with test r for subpopulation g. One minus each of these values yields the respective specificities and sensitivities of the two tests. Assuming conditional independence between the two test error rates, the multinomial probabilities associated with the cell frequencies are:

$P(A) = \pi_g (1-\beta_{1,g})(1-\beta_{2,g}) + (1-\pi_g)(\alpha_{1,g}\, \alpha_{2,g})$

$P(B) = \pi_g (\beta_{1,g})(1-\beta_{2,g}) + (1-\pi_g)(1-\alpha_{1,g})(\alpha_{2,g})$

$P(C) = \pi_g (1-\beta_{1,g})\, \beta_{2,g} + (1-\pi_g)(\alpha_{1,g})(1-\alpha_{2,g})$

$P(D) = \pi_g (\beta_{1,g}\, \beta_{2,g}) + (1-\pi_g)(1-\alpha_{1,g})(1-\alpha_{2,g})$

For a given subpopulation g, the data in Figure 1 provides only three independent cell frequencies from which to estimate the 5 parameters presented. Hence, the model is overparameterized for estimation purposes. However, by applying various assumptions about the parameters based on the reinterview methods discussed in the next section, the number of parameters are reduced sufficiently for estimation purposes.

3. Traditional Methods for Estimating Response Errors

In the response-bias type study by the use of preferred data collection techniques, the false positive and false negative rates, $\alpha_{2,g}$ and $\beta_{2,g}$ for the reinterview are presumed to be negligible. Therefore, the observed proportion of units classified by the reinterview as positive is an estimate of the true prevalence rate. From Figure 1, the estimate of the true prevalence rate can be computed from the observed frequency of the individuals classified in cells (A+B)/n. Similarly, the observed the value of (A+C)/n provides an estimate of the parent survey's estimate of the prevalence rate. The difference between these two values, defined as the net difference rate (NDR), estimates the response bias present in the parent survey data. This quality in the terms of the parameterization presented in section 2 is as follows.

$\hat{\pi}_g = P(A) + P(B)$

$NDR = P(C) - P(B)$

and $E(NDR) = -\pi_g\, \beta_{1,g} + (1-\pi_g)\, \alpha_{1,g}$

assuming $\alpha_{2,g} = 0$ and $\beta_{2,g} = 0$

Traditionally, the estimate of the net difference rate is the only estimate produced for response bias evaluations. Measures of the false positive and the false negative rates are not computed.

The measurement of response-variance, often referred to as the simple response variance (SRV) in terms of the parmeterization in section 2 is defined as follows,

$SRV = \pi_g\, \beta_{1,g}\, (1-\beta_{1,g}) + (1-\pi_g)\, \alpha_{1,g}\, (1-\alpha_{1,g})$

The SRV can be estimated from the response-bias type reinterview by noting that

$\alpha_{1,g} = P(C)/P(C+D)$ and $\beta_{1,g} = P(B)/P(A+B)$.

However, typically, a response-variance reinterview is conducted in such a fashion as to assume that $\alpha_{1,g} = \alpha_{2,g}$ and $\beta_{1,g} = \beta_{2,g}$. In this case the SRV is estimated from the sum of the observed proportions in cells B+C divided by 2. If a response-variance type reinterview is conducted, no measurement of the response bias is available under traditional methods.

4. Concerns with the Traditional Estimation Procedures

In a response-bias type reinterview, if the reinterview false positive and false negative rates are not negligible, the expected value of the estimates of the true prevalence rate and the NDR are as follows.

If $\alpha_{2,g}$ and/or $\beta_{2,g} > 0$. Then,

$E(\hat{\pi}_g) = \pi_g(1-\alpha_{2,g} - \beta_{2,g}) + \alpha_{2,g}$

$E(NDR) = P(C) - P(B) = -\pi_g\, (\beta_{1,g}-\beta_{2,g}) + (1-\pi_g)(\alpha_{1,g}-\alpha_{2,g})$.

Both differ from desired quantity by

$\pi_g(\alpha_{2,g} + \beta_{2,g}) - \alpha_{2,g}$

If in a response-variance type reinterview, the false-positive and false-negative rates are not identical to those from the parent survey, then the expected value of the estimate of the SRV is,

$E(SRV) = \pi_g/2\, (\beta_{1,g} + \beta_{2,g} - 2\, \beta_{1,g}\beta_{2,g}) +$

$(1-\pi_g)/2\, (\alpha_{1,g} + \alpha_{2,g} - 2\, \alpha_{1,g}\, \alpha_{2,g})$.

which differs from the desired result by

$\pi_g\, [(\beta_{1,g} - \beta_{2,g})(1/2 - \beta_{1,g})\,] +$

$(1-\pi_g)\, [(\alpha_{1,g} - \alpha_{2,g})(1/2 - \alpha_{1,g})]$.

The numerical impact of the biases presented above when the assumptions of the reinterview study design are not met can be quite varied depending on the size of the prevalence rates. For example, assume that the false-positive and the false-negative rate from the parent survey are both equal to 2% and that a reinterview-bias study is conducted which reduces these errors to 1%. For prevalence rates around a 50% level the impact of 1% error rate in the reinterview procedures has little impact on the estimate of π. However, if the true value of π is less than 10%, a substantial bias (almost 1%) will result from the same 1% error rate.

Hence, the results from the reinterview survey may be far from optimal if the reinterview objectives are not fully met. Given these concerns, we present the Hui and Walter methods as a means for evaluating the effectiveness of the reinterview procedures.

5. The Hui and Walter Method

The Hui and Walter methods were developed for the evaluation of diagnostic tests. Specific estimation procedures are presented in Hui and Walter's paper when two tests are applied simultaneously to each individual from S populations. Conditional on the true disease state, the errors from both tests are assumed independent. Given these factors, the observations are distributed as S independent multinomials whose likelihood includes S products of the probabilities presented in section 2.

Rather than making the assumption that one of the testing procedures will provide for error-free results which is comparable to the response-bias type study procedures, Hui and Walter assume that $\beta_{r,1} = \beta_{r,2}$ and $\alpha_{r,1} = \alpha_{r,2}$, i.e. the test errors are equal for all populations, but the prevalence rates differ by population. With this assumption, the number of parameters is reduced and estimation of the classification errors for both of the testing procedures is possible for S=2 subpopulations. The estimators and the variances for the estimators are provided in Hui and Walter's paper. Once the estimated variances for the parameters are obtained, estimated variances for the estimates of the NDR and the SRV can be developed (first order Taylor series approximations are presented in this paper).

The assumption of equal error rates across populations is easily justified for many diagnostic tests which should exhibit the same error levels across social and economic groups. In the survey environment, the prevalence rates may be highly correlated with the testing errors, therefore, a careful selection of the populations is needed to ensure the proper application of these methods.

6. The Content Reinterview Survey Program

The Content Reinterview Survey (CRS) is conducted following the Decennial Census to measure the response error associated with selected population and housing data items from the 1990 Decennial Census long-form sample questionnaire.

A variety of questionnaire items from the long-form sample questionnaire were selected for evaluation in the 1990 CRS. Some of these questions received a response-bias type study reinterview procedure and others a response-variance study method. Detailed probing questions were used to obtain responses for the response-bias type study procedures while the identical question format was used to obtain data for the response-variance type studies. Interviews for the CRS were conducted using Computer-Assisted Telephone Interviewing (CATI) with personal visit follow-up.

The 1990 CRS sample design consisted of a nationally representative sample of all housing units that received a long form questionnaire in the 1990 Decennial Census. Approximately 1 in 6 of the estimated 96 million housing units in the nation received a long-form questionnaire. From this universe a final systematic simple random sample of 12,891 census non-vacant housing units was selected.

7. The Census Questionnaire Items Selected for Evaluation in the CRS

The items which will be evaluated in this report will be limited to Spanish origin and employment status. The question and response categories for Spanish origin are presented on the next page. This question was evaluated using a response-variance study design, hence the question format was identical for the 1990 Census and the 1990 CRS.

"Is ... of Spanish/Hispanic Origin? For example: Mexican, Mexican-American, Chicano, Puerto Rican, Cuban, Spaniard, or from the Spanish-speaking countries of Central or South America."

1. No (not Spanish/Hispanic)
2. Yes, Mexican, Mexican American, Chicano
3. Yes, Puerto Rican
4. Yes, Cuban
5. Yes, Other Spanish/Hispanic

For this evaluation we collapsed the five categories into 1) Yes (responses #'s 2 - 5) and 2) No (response #1). The value of π_g was defined for this question as the prevalence rate of all persons defined as being of Spanish origin.

Employment status was evaluated with a response-bias type study design which used an additional probe to enhance the singular 1990 Census question. The census and the CRS format are provided below.

1990 Census:

"Last Year (1989) did ... work even for a few days, at a paid job or in a business or farm?"

1. Yes 2. No

1990 CRS:

"Did ... work at a job or in a business or on a farm at any time during 1989?"

1. Yes - Skip next question 2. No

"Did ... do any temporary, part-time, or seasonal work even for a few days during 1989?"

1. Yes 2. No

The two answers were recoded to a singular response comparable to the census data. For example, if a "Yes" response was obtained for the second question with a combination of a "No" response to the first, the recode was coded as a comparable "Yes" response.

To apply the Hui and Walter approach to this data we had to select two subpopulations which would exhibit different prevalence rates but equivalent classification errors. Given that rate of males and females by Spanish origin or employment status are not equal, and that both sexes as a whole should have the same level of classification errors, these two subgroups, males and females, should satisfy our conditions. The person data were restricted to cases whose sex responses agreed in both the 1990 Census and the 1990 CRS to eliminate any possible error due to the matching of the CRS to the census data.

8. Results of the CRS Analysis Using The Hui and Walter Methods

The results from the Hui and Walter estimation procedures are presented for Spanish origin and employment status in Table 1 on page 6.

For the Spanish origin question, the false positive rate point estimates are zero (+ or - 2%) for both the census and the CRS indicating that the question never classifies the persons as Spanish/Hispanic origin when they are not. Given this fact, the Hui and Walter model was reduced to only include the false positive rates and the prevalence rate parameters. (standard errors under the full model with the zero estimates and the reduced model are presented). In the reduced model, the estimated false negative rates are statistically significant from zero. This indicates that the question tends to classify Hispanic persons as not Hispanic. Note that the CRS false positive rate was also expected to be equal to the census values due to the use of a response-variance study method, but that the reinterview estimate is significantly lower than the census false positive rate with the point estimate for the reinterview at about one half the value for the census.

For the employment status question, the CRS reinterview procedures showed a reduction in the point estimates of the false negative and positive rate from the census of 15% to 40%, respectively. Both the CRS error rate estimates (α_2 =.116 and β_2 = .059) were statistically significant from zero indicating that the probing question which was designed to assist in measuring employment status was not sufficient in providing for completely accurate responses.

9. Properties of the Hui and Walter Methods

In analysis of the CRS data we have assumed that the test error rates are equal for the two populations studied, males and females. While this assumption seems reasonable in this study, the data structures studied do not allow for formal tests to be conducted to verify this assumption. Since the Hui and Walter method yields bias estimates when the classification errors differ between the two populations selected, we need to assess how sensitive the Hui and Walter methods are to breakdown in these data requirements relative to the level of bias in the traditional methods. While the answer to this question depends on several factors including the level of the prevalence rates and the classification errors which cannot be addressed fully in this report a few examples are presented for comparison in Table 2 on page 6.

In all of the examples in Table 2 we have assumed that the parent survey's false positive and false negative rate for population #1 are equal at a value of 5% ($\alpha_1 = \beta_1 = .05$) and that the reinterview reduces these values to 2% ($\alpha_2 = \beta_2 = .02$). In example #1, the error rates for the second population are equal to the first; therefore, the Hui and Walter estimate of π_1 is unbiased.

When Hui and Walter assumptions are violated, the effects on the resulting estimators are less detrimental when the difference between the prevalence rates for the two populations increases as indicated by a lower bias in the Hui and Walter estimates in examples 4 and 5 compared to the bias in examples 2 and 3. The Hui and Walter methods perform better when the error rates in the second population are not consistently higher or lower than the first as noted by the difference between examples 2 and 3 and the between examples 4 and 5 (In examples 3 and 5 the false negative rate is higher in the second population but the false positive rate is lower). Note that in example 5, the Hui and Walter estimate of the true .20 prevalence rate only incurs a bias of 0.0013 whereas the response bias assumptions impose a bias of 0.012.

The results in example 6 show that the Hui and Walter methods can tolerate even higher levels of violations in the assumptions when the prevalence rates drop below 10% before they provide equivalent levels of bias as those present in corresponding response-bias methods.

In summary, the Hui and Walter methods can provide substantially better estimates of the prevalence rate than the response-bias methods even when the required assumptions are violated providing the degree of the problems are not substantial. If the two populations studied under the Hui and Walter method have prevalence rates which are quite different in their relative size, even fairly large differences between the error rates in the two populations can be tolerated while still providing for an improved estimate in lieu of the traditional procedures.

10. Summary

In this paper we have shown that the Hui and Walter methods can be used to evaluate reinterview performance or provide for improved estimates of the prevalence rates for selected subpopulations.
In this study, the ability of the CRS reinterview methods to met either the response-bias study or response-variance study objectives appear to be somewhat ineffective. Given these results and that the Hui and Walter methods can only be used to develop estimates for certain subpopulations, continued research is warranted to develop other estimation methods or alternative reinterview design procedures which do not rely on presumed properties of the data for unbiased estimation.

References

Hui, S. L. and Walter, S. D. (1980). Estimating the error rates of diagnostic tests. Biometrics 36, 167-171

U.S. Bureau of the Census, (1985). Evaluating Censuses of Population and Housing, Statistical Training Document, ISP-TR-5, Washington D.C.

U.S. Bureau of the Census, (1986). Content Reinterview Study: Accuracy of Data for Selected Population and Housing Characteristics as Measured by Reinterview, Washington D.C

Acknowledgements

1. This paper reports 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.

2. Research conducted by Dr. Joseph Gastwirth, George Washington University is partially funded by a grant from the National Science Foundation (NSF).

TABLE 1

RESULTS OF HUI AND WALTER ANALYSIS OF 1990 CRS REINTERVIEW

Prevalence/ Question Evaluated and Study Method	POP	Sample Size Matched Census-CRS person records	Hui and Walter Model Results standard errors of the estimates in ()'s					Est. of Parent Survey π	Est. of π from Reint.	Response Error Measure
			Parent Survey Errors		Reinterview Survey Errors		π			
			α_1	β_1	α_2	β_2				
π = proportion of persons classified as being of **Spanish Origin** Response-Variance Type Study	Males	11143	0.000 (0.008)	0.135 (0.132) (.008)*	0.000 (0.012)	0.064 (0.099) (.006)*	0.082 (0.015) (.003)*	0.071 (.002)	0.077 (.003)	Reint. SRV .008 (.001) H&W SRV .010 (.009) (.001)*
	Females	12124					0.073 (0.015) (.002)*	0.063 (.002)	0.069 (.003)	Reint. SRV .006 (.001) H&W SRV .009 (.008) (.001)*
π = proportion of person classified as **Working in 1989** Response-Bias Type Study	Males	7348	0.134 (0.035)	0.096 (0.010)	0.116 (0.044)	0.059 (0.009)	0.854 (0.013)	0.792 (.005)	.821 (.005)	Reint. NDR -.029 (.004) H&W NDR .062 (.014)
	Female	7179					0.755 (0.017)	0.716 (.005)	.739 (.005)	Reint. NDR -.023 (.004) H&W NDR -.039 (.016)

* = Reduced model standard errors

TABLE 2

COMPARISON OF HUI AND WALTER ESTIMATES TO RESPONSE-BIAS TYPE REINTERVIEW ESTIMATES IN PRESENCE OF ERRONEOUS ASSUMPTIONS

EXAMPLE #	True π Pop 1	True π Pop 2	True Error Rates						Hui and Walter Est. of π Pop 1	Resp.-Bias Type Reinter-view Est. of π Pop 1	Bias in Hui and Walter Est. of π Pop 1	Bias in Response - Bias Type Reinter-view Est. of π POP 1
			α_1 and β_1 Pop 1	α_2 and β_2 Pop 1	α_1 Pop 2	α_2 Pop 2	β_1 Pop 2	β_2 Pop 2				
1.	.2	.4	.05	.02	.05	.020	.050	.020	.2000	.212	.0000	.012
2.	.2	.4	.05	.02	.060	.024	.060	.024	.2125	.212	.0125	.012
3.	.2	.4	.05	.02	.060	.024	.040	.016	.2023	.212	.0023	.012
4.	.2	.6	.05	.02	.060	.024	.060	.024	.2062	.212	.0062	.012
5.	.2	.6	.05	.02	.060	.024	.040	.016	.1988	.212	.0013	.012
6.	.09	.07	.05	.02	.065	.026	.065	.026	.0777	.1064	.0123	.0164