EVALUATING CENSUS DATA QUALITY USING INTENSIVE REINTERVIEWS: A COMPARISON OF U.S. CENSUS METHODS AND RASCH METHODS

Robert A. Johnson and Henry F. Woltman
U.S. Bureau of the Census

I. INTRODUCTION

This paper compares two statistical approaches to evaluating the quality of survey data obtained using different measurement procedures to measure the same phenomenon. The two approaches are the conventional U.S. Census approach (Hansen, et.al., 1961), which is also called "response bias estimation", and the Rasch latent trait approach (Rasch, 1960/ 1980). Both approaches require at least two repeated measures of the same variable. In order to estimate response bias, the Census approach regards one of the repeated measures as a standard of unbiasedness. An "intensive reinterview", i.e. a reinterview using detailed, probing questions that are personally addressed to the person whose characteristics are being measured, is often used as the standard.

This paper applies both approaches to data collected using what we call the "intensive reinterview design", the survey design used by the Census Bureau, in the 1976 National Content Test (NCT), to evaluate alternative question versions that were proposed for the 1980 Decennial Census. This design, which is also the basis for the 1986 NCT (currently in the field), has three features: 1) random assignment of households to (mail out/ mail back) questionnaire panels; 2) an intensive reinterview on the same topics (about two months after the initial interview in the NCT); and 3) independence of response errors in repeated measurements of the same units. The last assumption, analogous to the Rasch assumption of "local independence" (discussed in the following), is needed to assure that bias from the initial interview does not contaminate the reinterview standard of unbiasedness. Due to memory effects, the validity of this assumption may be poorly controlled by design. Moreover, differences between the initial interview and reinterview data might result from changes in the population rather than from any difference in the measurement procedures.

Section II compares the assumptions of the two approaches. Section III compares the results of applying the approaches to two binary test items, "work limitation" and "work prevention" due to disability, from the 1976 NCT. Section IV concludes that the two approaches differ fundamentally in their assumptions and implications for data content evaluations. Section V provides a short bibliography of each approach.

II. ASSUMPTIONS

Display 1 of Table 1 defines symbols for the observed cell frequencies of the "standard tables" from the intensive reinterview design that are utilized by both approaches. Within each of two experimental questionnaire panels, Panel A and Panel B, initial interview responses are cross-classified by the reinterview responses of the same respondents. Using the symbols a, b, c, and d to denote the cells of each 2 x 2 standard table is a convention in Census Bureau reports (e.g., U.S. Bureau of the Census, 1985). This convention assumes that each initial interview question version and the reinterview question version have only two response categories, denoted "Yes" and "No" in Table 1. We shall show that the assumptions of the Census and Rasch approaches lead to different parametrizations of the expected cell frequencies of the standard tables.

Consider first the conventional census approach. This approach arises from the "fixed bias version" (e.g. Lessler, 1984, pp. 413-414) of a general model for the decomposition of total survey mean-square error (Hansen, et.al., 1961; Bailar, 1976). Let Q = the true population proportion having the characteristic; let q = the sample proportion responding "yes" in one of the experimental panels (either Panel A or Panel B); and let q_r = the sample proportion in the reinterview.

Assuming that q_r is free of response bias and that the sample design (including any coverage improvement or nonresponse reduction procedures) is such that sample proportions are unbiased estimators of population proportions, the mathematics of expectation imply

$$MSE(q) = SV(q) + [E(q) - Q]^2, \qquad (1)$$

where MSE(q) is the mean-square error of q, SV(q) is the sampling variance of q, and the expectation is taken over all possible samples given the sample design. The last term is the squared response bias. It follows that an unbiased estimator of the response bias is the net difference rate (NDR):

$$\begin{aligned} NDR &= q - q_r \\ &= (a + c)/N - (a + b)/N \\ &= (c - b)/N \\ &= R_2p - R_1n, \qquad (2) \end{aligned}$$

where n = b/(a + b) is the false negative rate, i.e. the estimated proportion of those who have the characteristic, according to the reinterview, who were misclassified in the initial interview, p = c/(c + d) is the false positive rate, i.e. the estimated proportion of those who do not have the characteristic, according to the reinterview, who were misclassified in the initial interview, and R_1 and $R_2 = (1 - R_1)$ are the proportions responding "Yes" and "No", respectively, in the reinterview. In practice, biases due to nonresponse and incomplete coverage are likely to be present, but, provided these biases are constant across panels, comparing NDR's still gauges the relative response biases of the alternative question versions.

Generally, the Census approach assumes that each person in the population has a "true state" that corresponds to one of the response categories of a question. This seems more reasonable for some types of questions, e.g. male vs. female, than for others, e.g. prevented from working vs. not prevented from working. In contrast, the Rasch approach assumes that persons vary on an underlying continuum called the "latent trait". For example, persons might vary continuously between "not at all limited in working" and "totally prevented from working." Lessler (1984) points out that the existence of true states (by implication, the concept of response bias) has often been questioned when the phenomenon being measured seems "subjective" or "sensitive".

We think the Rasch approach clarifies the motivation for questioning the assumption in such cases: Assuming that the characteristic varies continuously and that the question is sensitive, slightly varying the wording or sequencing might change the meaning of the question, i.e. the restrictiveness of the implicit definition of the characteristic. If so, it is unreasonable to inquire which of two question versions is least biased because the two versions might yield different population proportions responding "Yes" but still be equally accurate, in the sense of discriminating among persons who have more or less of the characteristic. In the language of Rasch, this occurs if the item threshold, i.e. the level of the underlying latent trait that is required, on average, to produce a "Yes" answer, is higher for one question version than for the other.

Like Hansen, Rasch sought to separate the effects of the measurement procedure from the effects of the sampling procedure. However, rather than partitioning total error into measurement and sampling components, Rasch aimed to construct the measuring instruments in such a way that their properties could be analytically separated from the characteristics of particular samples. Rasch's concept of "measurement objectivity" specifies that the item threshold of a measuring instrument should be invariant when the instrument is applied to different subsamples of subjects. For example, a yardstick would be seriously impaired (would lack "objectivity") if it measured differently depending upon whether a rug, a table, or a picture was being measured (L.L. Thurstone, cited in Duncan, 1984).

Rasch claimed that two properties of measuring instruments, logistic form and local independence, were necessary and sufficient for measurement objectivity. Rasch proved the sufficiency of these two conditions; their necessity was proved by Douglas and Wright (1986). Consider either Panel A or Panel B and let p_{ji} = the probability that the jth respondent in the panel gives a "Yes" response in the initial interview; let p_{jr} = the probability that the jth respondent gives a "Yes" response in the reinterview; and let t_j = the value of a positive random variable, the unobserved latent trait value of the jth respondent (assumed unchanged between interviews).

The assumption of logistic form states that

$$p_{ji} = it_j/(1 + it_j)$$
$$\text{and } p_{jr} = rt_j/(1 + rt_j) \qquad (3)$$

where i and r are the item parameters for the initial interview and reinterview measurements respectively. (Equivalently, the logit of the response probability equals the sum of a subject parameter and an item parameter.) The item threshold of a question is simply the reciprocal of the corresponding item parameter in (3). Model (3) is also called the simple (one parameter) logistic model for binary responses (e.g. Andersen, 1980, Chap. 6).

The assumption of local independence states that the initial interview and reinterview probabilities for the jth respondent can be multiplied to get interview-reinterview joint probabilities. That is, the probabilities that the jth respondent is tabulated in cells a, b, c, and d of the standard table are:

$$\begin{aligned} p_{ja} &= irt_j^2/[(1 + it_j)(1 + rt_j)], \\ p_{jb} &= rt_j/[(1 + it_j)(1 + rt_j)], \\ p_{jc} &= it_j/[(1 + it_j)(1 + rt_j)], \\ p_{jd} &= 1/[(1 + it_j)(1 + rt_j)]. \end{aligned} \qquad (4)$$

The expected frequencies are obtained by summing each probability in (4) over the sample. In Panel A, we sum each probability from $j = 1$ to $j = N_A$:

$$\begin{aligned} E(a) &= irA_2; \\ E(b) &= rA_1; \\ E(c) &= iA_1; \\ E(d) &= A_0 \end{aligned} \qquad (5)$$

A_2, A_1, and A_0 denote sums of terms involving the t_j's and are called the "composition parameters" for two "Yes" answers, one "Yes" answer, and no "Yes" answers respectively. In Panel B, denote these parameters using B_2, B_1, and B_0.

Clearly, the expected cell frequencies of (5) satisfy measurement objectivity because the item parameter of the initial interview instrument, relative to the reinterview item parameter (assumed constant across panels), can be estimated using $E(c)/E(b) = i/r$, which does not depend on the values of the t_j's in the particular sample. In the Rasch approach, the invariance of the ratio $E(c)/E(b)$ across diverse subpopulations becomes a key empirical question.

Duncan (1984) shows that the composition parameters generally measure the effects of sample heterogeneity with respect to the latent trait. Consider the usual measure of association, the odds-ratio, applied to the fourfold table in (5): $[E(a)E(d)]/[E(b)E(c)] = A_2A_0/A_1^2$. Given the Rasch parametrization, then, the odds-ratio is strictly a function of the composition parameters. If we set $t_j = t$, a constant, for all j, then the odds-ratio equals unity (Eqs. (4)-(5)), implying that there exists no association between the interview and reinterview measurements. In the language of the Census approach, it follows that the reliability (response consistency) of measurement is no greater than expected from chance. Hence, while the Census approach interprets reliability (for binary data, the inverse of what is called "response variance") as a property of the measuring instrument, the Rasch approach interprets reliability as gauging heterogeneity of the population on the unobserved latent trait.

When subjects are randomly assigned to panels, as in the intensive reinterview design, the distribution of the latent trait can be assumed to be the same, except for random error, in each of the experimental panels. Applied to (5), this implies that the composition parameters differ across panels only as a function of panel differences in the experimental item parameter. Hence, if panel item parameters are inferred to be equal, random assignment makes possible a test of the Rasch assumptions of logistic form and local independence (Section III).

Rather than summing over the sample, as in (5), Cressie and Holland (1983) integrate the joint probabilities against the density of the latent trait in the population. Using this approach, they prove that the composition parameters are the moments of a positive random variable. For each 2 x 2 table of Table 1, Display 1, this implies that the odds-ratio must be greater than or equal to unity. Note that these conditions are satisfied for the data of Table 2. Display 2. When more than two repeated measures or more than two response categories are available, the moment inequalities imply "isotropy conditions" (Hout, et.al., 1986, following G.U. Yule) that must be satisfied for a unidimensional Rasch model to be tenable.

Displays 2 and 3 of Table 1 show the Census and Rasch parametrizations, respectively, of expected cell frequencies, assuming multinomial sampling and the intensive reinterview design. The Census parametrization has one degree of freedom because random assignment to panels and the common reinterview procedure (i.e. the same reinterview procedure being used in each panel) imply that the reinterview marginal proportions R_1 and R_2 are constant across panels, except for random error. Note that the response bias is not a parameter. Indeed, the NDR is not a sufficient statistic for any parameter of any elementary discrete sampling model, nor for complex sampling extensions of such basic models (e.g. Cox, 1970, Chap. 2). The Rasch parametrization has zero degrees of freedom, but, significantly, the assumption of a common reinterview implies that r is constant across panels (hence, is not subscripted in Table 1, Display 3). As shown in III, this makes it possible to estimate the panel item parameters i_A and i_B.

III. EVALUATION OF WORK DISABILITY ITEMS

Display 1 of Table 2 shows Panel A and Panel B question versions for measuring work limitation and work prevention from the 1976 NCT. Display 2 of Table 2 shows standard tables for two binary response variables, "limits or prevents work" and "prevents work", that were constructed using these data. General rules for partitioning the likelihood-ratio chi-square statistic (Goodman, 1969) were applied to each of the panel comparisons of Table 2. For each binary variable, the chi-square for the total difference between Panels A and B (3 df) was decomposed into three independent chi-square components (1 df each), using first the Census parametrization (Table 1, Display 2) and then the Rasch parametrization (Table 1, Display 3):

Total panel difference: Apply the usual chi-square test of row-column independence to the 2 x 4 table (3 df):

$$T = \begin{vmatrix} a_A & b_A & c_A & d_A \\ a_B & b_B & c_B & d_B \end{vmatrix}$$

Census decomposition: Apply the usual chi-square test to each of three 2 x 2 tables (1 df each):

$$C1= \begin{vmatrix} a_A+b_A & c_A+d_A \\ a_B+b_B & c_B+d_B \end{vmatrix} \quad C2= \begin{vmatrix} a_A & b_A \\ a_B & b_B \end{vmatrix} \quad C3= \begin{vmatrix} c_A & d_A \\ c_B & d_B \end{vmatrix}$$

C1 tests the equality of reinterview margins (which should not be rejected, given the design), C2 tests the equality of false negative rates, and C3 tests the equality of false positive rates.

Rasch decomposition: Apply the usual chi-square test to each of three 2 x 2 tables (1 degree of freedom each):

$$R1= \begin{vmatrix} b_A & c_A \\ b_B & c_B \end{vmatrix} \quad R2= \begin{vmatrix} b_A+c_A & a_A+d_A \\ b_B+c_B & a_B+d_B \end{vmatrix} \quad R3= \begin{vmatrix} a_A & d_A \\ a_B & d_B \end{vmatrix}$$

R1 tests the equality of item parameters, R2 tests the equality of inconsistency ratios, and R3 tests the equality of the distributions of consistently classified cases. Generally, the sum (R2 + R3) tests the hypothesis that the composition parameters are invariant across panels. But, given random assignment to panels, if the first hypothesis, $i_A = i_B$, is not rejected, then the sum (R2 + R3) tests the Rasch assumptions.

Display 1 of Table 3 shows the results of the alternative decompositions. With regard to "limits or prevents", neither approach can have a basis for preferring either panel since the chi-square for the total panel difference (3.0, 3 df) is not significant. Given the randomization and that R1 is not significant, the sum (R2 + R3) tests the Rasch assumptions (X^2 = 1.1, 1 df). With regard to "prevents", the Census approach infers a panel difference in the false positive rate (C3 = 7.1, 1 df), while the Rasch approach infers a panel difference in the item threshold (R1= 7.0, 1 df).

Display 2 of Table 3 shows the parameter estimates corresponding to the chi-square components of Display 1 and their standard errors assuming simple random sampling. Although the NDR is not a sufficient statistic, the Census approach still infers (e.g. using a two-sample t-test, with a pooled variance estimate computed from the standard errors) that the Panel A version of "prevents" is more biased than the Panel B version, since the NDR in Panel A (0.89) is further from zero than the NDR in Panel B (-0.24). The Rasch approach infers that the item threshold of Panel A (-0.6, on the log scale) is lower than the item threshold of Panel B (0.2, on the log scale).

It is not surprising that the Census approach tends to find a significant difference in the NDR whenever the Rasch approach finds a significant difference in the item threshold, since the two measures are mathematically closely related (Table 3, Display 2, column headings). The main difference is that, while the Census approach refers each NDR to a standard of zero bias, the Rasch approach compares the raw magnitudes of the item thresholds. Since, unlike the NDR, the item threshold is scale-free, setting r = 1 in in computing i_A and i_B (in each panel, given the common reinterview) is justified in Table 3, Display 2.

The differences between the Census and Rasch approaches may not be fully evident from the binary response case of Tables 1-3. Unlike the Census model, the Rasch model can be readily extended to the evaluation of ordinal, nonnumeric data. Indeed, since it assumes that the trait varies continuously, the Rasch approach encourages collecting data on as many "degrees" or "levels" as are affordable. For example, we might hypothesize that the categories "not limited", "limited", and "prevented" comprise what Masters and Wright (1984) call a "unidimensional rating scale model". However, Hout, <u>et.al.</u> (1986) argue that, when three or more ordered categories of response or more than two repeated measures are available, the existence of multiple dimensions of meaning of a question is a central issue. Given ordered polytomous responses, they show how it is sometimes possible, depending upon the magnitudes of population correlations among the dimensions, to empirically distinguish multidimensional from unidimensional Rasch models. For example, in measuring disability, two dimensions of meaning might be 1) the medical degree of limitation and 2) the degree of sensitivity to the labels "limited" and "prevented".

To make progress with Rasch, one also needs to investigate departures from measurement objectivity by testing the invariance of the ratio c/b across subclasses of nonexperimental variables that are thought to be strongly related to the trait being measured or to the way a question is interpreted (Duncan, 1984), e.g. medical condition and type of work in evaluating work disability items. Potentially, one might be able to untangle the different ways that respondents interpret a question.

IV. CONCLUSIONS

1.) The Census and Rasch approaches differ in their assumptions about i) the relationship of the fixed response categories of a survey item to the phenomenon being measured, ii) the type of response format to be analyzed, and iii) the class of probability distributions, continuous vs. discrete, that is appropriate for modeling the data: The <u>Census</u> approach assumes that respondents have true states that correspond to the response categories of a question. This implies that no information is lost if each response category is analyzed separately by dichotomizing the data as in Section III. The net difference rate (NDR) seems appropriate for normally distributed data, but the NDR is not a sufficient statistic for discrete sampling models. The <u>Rasch</u> approach assumes that respondents have values on an underlying continuum called the latent trait and that a respondent's probability of a giving a particular answer to an item measuring the trait is determined, via the logistic function, by that respondent's trait value. The assumptions of logistic form and local independence follow from the prescription of measurement objectivity and lead to reinterpretations of two conventional census evaluation concepts, response bias (reinterpreted as a difference in item thresholds) and response consistency (reinterpreted as depending upon the heterogeneity of the population being measured as well as upon the particular measuring instrument). The Rasch approach favors analyzing the original response format of a question (preferably, more than two ordered responses and/or more than two repeated measures) rather than any collapsed version of the data. The Rasch approach always estimates parameters of elementary discrete sampling models.

2.) The Rasch model is more readily tested than the Census model. Applied to contingency table data, the Rasch assumptions imply i) a specific loglinear model called quasi-symmetry (Bishop, Fienberg, and Holland, 1975, Chap. 8), ii) moment inequalities called "isotropy conditions" (Cressie and Holland, 1983; Hout, Duncan, and Sobel, 1986), and iii) invariance of item parameters across subclasses of non-experimental control variables (Duncan, 1984). In the two evaluations of this paper, each restricted to two measures of a binary response variable, quasi-symmetry has zero degrees of freedom (hence cannot be tested), the isotropy conditions are satisfied, and tests of invariance of meaning across subclasses are not applied. As shown in III, random assignment to panels provides an additional means of testing the Rasch assumptions in the case that panel item parameters are inferred to be equal. The Census assumption of an unbiased reinterview, on the other hand, cannot be tested unless one has available a third set of measurements for the same subjects that can be assumed to be <u>still</u> more accurate.

3.) In the Rasch approach, three criteria of good measurement supplant the Census criterion of unbiasedness: i) the degrees to which different question versions measure purely what they are intended to measure (unidimensionality), ii) the costs and benefits of different item thresholds for data uses and iii) the degrees of stability of the item parameters of different question versions (invariance of "meaning"). These criteria seem more complex to apply than the Census criterion of unbiasedness.

V. REFERENCES

CONVENTIONAL CENSUS APPROACH

BAILAR, B.A., 1976. Some sources of error and their effects on census statistics, <u>Demography</u> 13: 273-286.

HANSEN, M.H., W.N. HURWITZ, and M. BERSHAD, 1961. Measurement errors in censuses and surveys, <u>Bull.Int.Statist.Inst.</u> 38,#2: 359-374.

LESSLER, J.T., 1984. Measurement error in surveys, pp. 405-440 in Turner, C.F. and E. Martin, eds., <u>Surveying Subjective Phenomena</u>, V. 2, New York: Russell Sage.

U.S. BUREAU OF THE CENSUS, 1985. <u>Evaluating Censuses of Population and Housing</u>, STD ISP-TR-5, Washington, D.C.

RASCH APPROACH

CRESSIE, N. and P. HOLLAND, 1983. Characterizing the manifest probabilities of latent trait models, <u>Psychometrika</u> 48: 129-141.

DOUGLAS, G. and B. WRIGHT, 1986. The two category model for objective measurement, MESA #34, Dept. of Ed., U. of Chicago.

DUNCAN, O.D., 1984. The latent trait approach in survey research, pp. 210-227 in Turner, C.F. and E. Martin, eds., <u>Surveying Subjective Phenomena</u>, V. 1, New York: Russell Sage Foundation.

HOUT, M., O.D. DUNCAN, and M. SOBEL, 1986. Association and heterogeneity. forthcoming in Clogg, C.C., ed., <u>Sociological Methodology 1986</u>, San Francisco: Jossey-Bass.

MASTERS, G. and B. WRIGHT, 1984. Essential process in a family of measurement models, <u>Psychometrika</u> 49: 529-544.

RASCH, G., 1960/1980. <u>Probabilistic Models for Some Intelligence and Attainment Tests</u>, 2nd Ed., U. of Chicago.

OTHER REFERENCES

ANDERSEN, E.B., 1980. <u>Discrete Statistical Models with Social Science Applications</u>, Amsterdam: Noth-Holland.

BISHOP, Y., S. FIENBERG, and P. HOLLAND, 1975. <u>Discrete Multivariate Analysis, Theory and Practice</u>, Cambridge: M.I.T.

COX, D.R., 1970. <u>Analysis of Binary Data</u>, London: Methuen.

FIENBERG, S.E., 1980. <u>The Analysis of Cross-Classified Categorical Data</u>, 2nd ed., Cambridge, Mass.: M.I.T.

GOODMAN, L.A., 1969. "On partitioning X^2 in three-way contingency tables," <u>J.Roy.Statist.Soc. Ser. B</u> 31: 486-498.

Table 1, Display 1: General symbols for cross-classifications of initial interview and reinterview binary responses. Display 2: Census parametrization (1 df). Display 3: Rasch parametrization. Define $N_A = a_A+b_A+c_A+d_A$ = sample size of Panel A and $N_B = a_B+b_B+c_B+d_B$ = sample size of Panel B.

DISPLAY 1

	Initial Interview			
	Panel A		Panel B	
Reinterview	Yes	No	Yes	No
Yes	a_A	b_A	a_B	b_B
No	c_A	d_A	c_B	d_B

DISPLAY 2

	Initial Interview			
	Panel A		Panel B	
Reinterview	Yes	No	Yes	No
Yes	$N_A R_1(1-n_A)$	$N_A R_1 n_A$	$N_B R_1(1-n_B)$	$N_B R_1 n_B$
No	$N_A R_2 p_A$	$N_A R_2(1-p_A)$	$N_B R_2 p_B$	$N_B R_2(1-p_B)$

DISPLAY 3

	Initial Interview			
	Panel A		Panel B	
Reinterview	Yes	No	Yes	No
Yes	$i_A r A_2$	$r A_1$	$i_B r B_2$	$r B_1$
No	$i_A A_1$	A_0	$i_B B_1$	B_0

Table 2. Data on Work Limitation and Prevention, 1976 National Content Test. Display 1: Experimental Question Versions, Panel A and Panel B. Display 2: Cross-Classifications of Initial Interview and Reinterview Responses (Respondents Aged 18-64).

DISPLAY 1

Panel A: Does this person now have a physical, mental or other health condition or handicap which has lasted for 6 months or more and which...

	YES	NO
e. Limits the kind or amount of work this person can do..	___	___
f. Prevents this person from working at any job or business....................................	___	___

Panel B: Does this person now have a physical, mental or other health condition which has lasted for six months or more and which limits or prevents the following:

	Prevents	Limits but does not prevent	No limitation
e. Working at any job or business.......	___	___	___

DISPLAY 2

LIMITS OR PREVENTS WORK

Initial interview

Reinterview	Panel A Yes	Panel A No	Total		Panel B Yes	Panel B No	Total
Yes	140	84	224	Yes	158	94	252
No	86	3047	3133	No	71	3051	3122
Total	226	3131	3357	Total	229	3145	3374

PREVENTS WORK

Initial interview

Reinterview	Panel A Yes	Panel A No	Total		Panel B Yes	Panel B No	Total
Yes	84	37	121	Yes	83	48	131
No	67	3169	3236	No	40	3203	3243
Total	151	3206	3357	Total	123	3251	3394

Table 3. Display 1: Census and Rasch Decompositions of the Likelihood-Ratio Chi-Square for the Total Panel Difference. Display 2: Census and Rasch Parameter Estimates. (In parentheses, standard errors assuming simple random sampling.)

DISPLAY 1

DATA	TOTAL T (3 df)	CENSUS C1 (1 df)	C2 (1 df)	C3 (1 df)	RASCH R1 (1 df)	R2 (1 df)	R3 (1 df)	(R2 + R3) (2 df)
limits or prevents	3.0	1.6	0.0	1.4	1.9	0.1	1.0	1.1
prevents	8.5	0.4	1.0	7.1	7.0	1.5	0.0	1.5

DISPLAY 2

CENSUS PARAMETRIZATION

Data	Panel	percentage yes, reinterview $=100x(a+b)/N$ $=100xR_1/(R_1+R_2)$		false negative rate (n) $=100xb/(a+b)$		false positive rate (p) $=100xc/(c+d)$		net difference rate $=100x(c-b)/N$ $=100x(R_2p-R_1n)$	
limits	A	6.7	(0.4)	37.5	(3.2)	2.7	(0.3)	0.06	(0.39)
or prevents	B	7.5	(0.5)	37.3	(3.0)	2.3	(0.3)	-0.68	(0.38)
prevents	A	3.6	(0.3)	30.6	(4.2)	2.1	(0.3)	0.89	(0.30)
	B	3.9	(0.3)	36.6	(4.2)	1.2	(0.2)	-0.24	(0.28)

RASCH PARAMETRIZATION

Data	Panel	Base N (persons)	item threshold $=\ln(b/c)$[1]		consistency parameter $=\ln((a+d)/(b+c))$		distribution, consistent cases $=\ln(d/a)$	
limits	A	3357	0.0	(0.2)	2.9	(0.1)	3.1	(0.1)
or prevents	B	3374	0.3	(0.2)	3.0	(0.1)	3.0	(0.1)
prevents	A	3357	-0.6	(0.2)	3.4	(0.1)	3.6	(0.1)
	B	3374	0.2	(0.2)	3.6	(0.1)	3.7	(0.1)

1. Without loss of generality, we set $r = 1$ in computing the item thresholds. The item threshold is the natural logarithm of the reciprocal of the item parameter i in the Rasch parameterization of expected cell frequencies, Table 1, Display 3.