

T. Timothy Chen, The Upjohn Company

Abstract. Various randomized response schemes are considered as purposive misclassification models with known misclassification matrices. These matrices are the transition matrices from the true variables to the randomized response (misclassified or false) variables and are determined by the researcher through randomization devices. The observed data consist of randomized response variables, rather than their true variables. This paper develops a method of investigating association among the true variables by fitting log-linear models upon them and by using the observed data. The approach can be viewed from two perspectives: one uses the recursive system of log-linear models, the other explores the log-linear models based on flats.

Key Words: Randomized Response; Misclassification Models; Contingency Tables; Indirect Observation; Recursive System; Log-linear Models.

1. **Introduction.** A misclassification problem happens when we use an inaccurate instrument (which may be a physical apparatus, a survey questionnaire, or an interviewer) to measure a categorical variable. The observed value does not necessarily represent the individual's true value, and the structure of misclassification is unknown.

Suppose we are interested in only one variable j which has J possible categories; due to misclassification, we observe a false variable i with I possible categories. (We will use i, j , etc. to denote both the categorical variables and the indexes of their categories.) We can use a two-dimensional $I \times J$ contingency table to represent the situation: the first dimension is the misclassified or false variable and the second dimension is the true variable. Let the probability of any observation having (i, j) as its false and true classification be

π_{ij} ($\sum_{ij} \pi_{ij} = 1$). The elements $\{a_{i,j}\}$ of a

misclassification matrix A is defined as

$a_{i,j} = \pi_{ij} / \pi_{+j}$, which is the conditional

probability of any observation having i as the false classification given that it has j as the true classification. Chen (1979) discussed the unknown misclassification situation where the matrix A is unknown.

In human surveys, individuals possessing a stigmatizing trait may answer the sensitive question relating to this trait evasively. The data obtained are misclassified again with an unknown structure of misclassification. In order to overcome this problem, randomized response techniques are proposed to let the individuals apply a prespecified misclassification structure to their true status;

this protects the privacy of individuals and we can still make inferences about the population as a whole. Dalenius (1977) called the randomized response techniques as privacy transformations. In this case, the matrix A is prespecified with known numerical values.

For example, Warner (1965) proposes a RR scheme for $I, J = 2$ (yes or no). If a person is in "yes" category ($j=1$), he will answer "yes" ($i=1$) with a probability p and "no" ($i=2$) with a probability $1-p$. If a person is in "no" category, he will answer "yes" with a probability $1-p$ and "no" with a probability p . Thus the observed value i is a misclassified value, and the matrix A is

$$\begin{matrix} & j=1 & j=2 \\ i=1 & \begin{bmatrix} p & 1-p \end{bmatrix} \\ i=2 & \begin{bmatrix} 1-p & p \end{bmatrix} \end{matrix}, \quad (1.1)$$

where p is determined by the researcher.

Greenberg *et al.* (1969) also propose a scheme for $I, J = 2$ which involves an unrelated question with binary response (yes or no). A person selects the original sensitive question to answer with a probability p and the unrelated question to answer with a probability $1-p$. If the population proportion of "yes" to the unrelated question is known, say π_y , then the matrix A is

$$\begin{matrix} & j=1 & j=2 \\ i=1 & \begin{bmatrix} p + (1-p)\pi_y & (1-p)\pi_y \\ (1-p)(1-\pi_y) & p + (1-p)(1-\pi_y) \end{bmatrix} \\ i=2 & \end{matrix} \quad (1.2)$$

In a given applied problem, we will usually investigate many categorical variables simultaneously. If some of the variables are subject to an unknown misclassification structure, then a double sampling scheme can be used and Chen (1979) proposes first fitting an overall log-linear model to find out the misclassification structure and then fitting another log-linear model upon the distribution of the true variables. In the randomized response, we can apply a prespecified misclassification structure to some of the variables; hence, the overall log-linear model is known. We can then fit log-linear models upon the distribution of the true variables using only the randomized response data. This paper will explain a three-dimensional example in detail, and the extension to higher-dimension should be very straightforward. For details about the log-linear models, please see Haberman (1974a), Bishop, Fienberg, and Holland

(1975), and Fienberg (1977). For explanations about the randomized response techniques, please see Cochran (1977).

2. Independent Misclassification. We will begin with an independent misclassification model in a three-dimensional (X, Y, Z) contingency table. Suppose both questions pertaining to variables Y and Z are sensitive, and we decide to use separate RR scheme for each variable, thus we have an independent misclassification model. Let the false variables (observations from RR's) for Y and Z be the first and second dimensions, variable X be the third dimension, and true variables for Y and Z be the fourth and fifth dimensions of an overall five-dimensional contingency table. Our overall log-linear model for the misclassification structure is

$$\pi_{ijklm} = \pi_{++klm} \left(\frac{\pi_{i+++}}{\pi_{++++}} \right) \left(\frac{\pi_{+j++}}{\pi_{++++}} \right) \quad (2.1)$$

where $i=1, \dots, I$, $j=1, \dots, J$, $k=1, \dots, K$, $l=1, \dots, L$, and $m=1, \dots, M$. The last two terms are conditional probabilities which correspond to misclassification matrices of two RR schemes. These matrices have known numerical values. The equation (2.1) means we have a log-linear model $H_{(14,25,345)}$ for the cell probabilities $\{\pi_{ijklm}\}$. The terms inside the parenthesis under a H notation denotes the highest u-terms in a hierarchical log-linear model (see Chen and Feinberg, 1976, and Chen, 1979).

From our design we cannot observe l and m of the true Y and Z, instead we observe i and j of the false Y and Z plus k of the variable X for each individual. Consequently, we have c_{ijk} observed counts in the cell (i,j,k) of an observed three-dimensional table with the cell probability π_{ijk++} . (Let N be the total sample size, $N = \sum \sum c_{ijk}$).

Now we want to make the inference about the original cell probabilities $\{\pi_{++klm}\}$ based on the observed counts $\{c_{ijk}\}$. More specifically, we like to build log-linear models on $\{\pi_{++klm}\}$.

The models we build on $\{\pi_{++klm}\}$ will be denoted by H^* . For example, if we want to fit no second-order interaction model, then we denote it as $H^*(34,35,45)$. The model $H_{(14,24,345)}$ plus $H^*(34,35,45)$ is called a recursive system of log-linear models in Fienberg (1977). It was first explored by Goodman (1973) in a modified path analysis context to analyze multi-dimensional contingency tables with some variables posterior to others.

From the other perspective, if we take logarithms of both sides of equation (2.1), then we have

$$\log \pi_{ijklm} = \log \pi_{++klm} + \log \left(\frac{\pi_{i+++}}{\pi_{++++}} \right) \left(\frac{\pi_{+j++}}{\pi_{++++}} \right) \quad (2.2)$$

for all i, j, k, l, m .

Since the last term is fixed, our log-mean vector is in a flat, rather than a linear manifold. Thus, we have log-linear models based on flats as first discussed in Chapter 9 of Haberman (1974a). The observed data we have is called frequency tables derived by indirect observation in Haberman (1974b). Thus, this paper explores log-linear models based on flats for indirectly observed tables.

3. Log-linear Models. Continuing the discussion of the previous section, we will first try to estimate $\{\pi_{++klm}\}$ by the maximum likelihood under the overall model $H_{(14,25,345)}$ and $H^*(345)$. In order to eliminate the identifiability problem, we require the two misclassification matrices being of ranks L and M . The maximum likelihood equations are

$$\hat{\pi}_{++klm} = \sum_{ij} \left[\frac{c_{ijk}}{(N \hat{\pi}_{ijk++})} \right] \hat{\pi}_{ijklm} \quad (3.1)$$

for all k, l, m ,

and $\{\hat{\pi}_{ijklm}\}$ should satisfy the equation (2.1), i.e.,

$$\hat{\pi}_{ijklm} = \hat{\pi}_{++klm} \left(\frac{\hat{\pi}_{i+++}}{\hat{\pi}_{++++}} \right) \left(\frac{\hat{\pi}_{+j++}}{\hat{\pi}_{++++}} \right) \quad (3.2)$$

for all i, j, k, l, m .

The equation (3.1) is similar to the equation (2.3) in Haberman (1974b) except that the log-mean vector is in a flat, rather than a linear manifold. The right-hand sides of the equation (3.1) are the proportional allocation of c_{ijk}/N to the (i,j,k,l,m) cell according to $\{\hat{\pi}_{ijklm}\}$ and then sum over i and j . The maximum likelihood equations correspond to highest u-terms in the H^* model.

The solution of these maximum likelihood equations can be obtained by an iterative procedure as follows: Take

$$\pi_{++klm}^{(0)} = 1/KLM, \text{ for all } k, l, m, \text{ then}$$

$$\pi_{ijklm}^{(v)} = \pi_{++k1+}^{(v)} \left(\frac{\pi_{i++1+}}{\pi_{+++1+}} / \frac{\pi_{+j++m}}{\pi_{++++m}} \right), \quad (3.3)$$

and

$$\pi_{++k1+}^{(v+1)} = \sum_{ij} \left[\frac{c_{ijk}}{N\pi_{ijk++}^{(v)}} \right] \pi_{ijklm}^{(v)} \quad (3.4)$$

This kind of iterative scaling procedure is used very extensively in the contingency table literature, see Haberman (1975) and Chen and Fienberg (1976). After obtaining $\{\hat{\pi}_{ijklm}\}$, we can test goodness-of-fit between $\{c_{ijk}\}$ and $\{N\hat{\pi}_{ijk++}\}$ by a likelihood ratio or Pearson chi-square test with degrees of freedom $(I-J)(M-K)$. If the randomization devices are doing well, most likely the fit should be good also.

Sometimes, if the sample size is small, then $\{\hat{\pi}_{ijklm}\}$ could be at the boundary of the parameter space and the log-likelihood is not well defined at this point. In this case, we will say the maximum likelihood estimates of $\{\pi_{ijklm}\}$ do not exist, and the $\{\hat{\pi}_{ijklm}\}$ obtained through the iteration cannot be used in the goodness-of-fit tests.

If we want to fit the no second-order interaction log-linear model $H^*(34,35,45)$ for $\{\pi_{++k1+}\}$, the maximum likelihood equations are

$$\hat{\pi}_{++k1+} = \sum_{ijm} \left[\frac{c_{ijk}}{N\hat{\pi}_{ijk++}} \right] \hat{\pi}_{ijklm}, \quad (3.5)$$

$$\hat{\pi}_{++k+m} = \sum_{ijl} \left[\frac{c_{ijk}}{N\hat{\pi}_{ijk++}} \right] \hat{\pi}_{ijklm}, \quad (3.6)$$

and

$$\hat{\pi}_{+++1m} = \sum_{ijk} \left[\frac{c_{ijk}}{N\hat{\pi}_{ijk++}} \right] \hat{\pi}_{ijklm}, \quad (3.7)$$

for all $k, 1$, and m .

Here $\{\hat{\pi}_{ijklm}\}$ should again satisfy the equation (3.2). The solution can be obtained by starting with the same initial values. At the beginning of the v -th stage, we determine $\{\pi_{ijklm}^{(v)}\}$ by (3.3). Then we determine $\{\pi_{++k1+}^{(v+1)}\}$, $\{\pi_{++k+m}^{(v+1)}\}$, and $\{\pi_{+++1m}^{(v+1)}\}$ by (3.5), (3.6), and (3.7), and adjust $\{\pi_{++k1+}^{(v)}\}$ according to the model $H^*(34,35,45)$ to $\{\pi_{++k1+}^{(v+1)}\}$ which satisfy three two-dimensional marginal constraints. This adjustment will be carried out by a separate iterative scaling procedure. This

entire procedure is similar to the one given in Chen (1979), except here we do not need an iteration for the overall model $H^*(14,25,345)$. After obtaining $\{\hat{\pi}_{ijklm}\}$, we test the goodness-of-fit with degrees of freedom $(I-J)(M-K) + (K-1)(L-1)(M-1)$. For other simpler H^* models, we can similarly fit and test. Thus, we can choose a best model to describe the original cell probabilities $\{\pi_{++k1+}\}$.

In our example, since the value k is not misclassified, the result we obtained is also valid for the product multinomial sampling with the multinomial in each k category.

The similarity between our problem and the latent structure problem discussed in Goodman (1974a, b) and Haberman (1975) is that both have indirectly observed tables: the difference between the two is that our log-linear models are based on flats and the latent structure models are based on manifolds.

4. Equivalent Models Under Independent Misclassification. In early literature, Bross (1954), Rubin, Rosenbaum, and Cobb (1956), and Mote and Anderson (1962) investigated a special case of our example with $J=M=1$ and $I=L$, and Assakul and Proctor (1967) and Drane (1976) investigated another special case with $K=1$, $J=M$, and $I=L$. When J, M , or K equals 1, we will ignore that dimension. Under the independent misclassification model $H^*(14,34)$ and the rank of the misclassification matrix being L , the first group of authors have shown that the marginal independence hypothesis $H^*(3,4)$ is equivalent to the marginal independence hypothesis $H^*(1,3)$. Therefore, the test of the model $H^*(3,4)$ is the test of $H^*(1,3)$ on the observed data. Under the independent misclassification model $H^*(14,25,45)$ and the ranks of the transformation matrices being L and M , the second group of authors have shown that the marginal independence model $H^*(4,5)$ is equivalent to the marginal independence model $H^*(1,2)$. Again, the test of the model $H^*(4,5)$ is the test of $H^*(1,2)$ on the observed data.

In our example, if $J=M$ and $I=L$, under the model $H^*(14,25,345)$ and the ranks of misclassification matrices being L and M , the models $H^*(34,35)$, $H^*(34,5)$, $H^*(4,35)$, $H^*(3,45)$ and $H^*(3,4,5)$ are equivalent to the models $H^*(13,23)$, $H^*(13,2)$, $H^*(1,23)$, $H^*(3,12)$, and $H^*(1,2,3)$, respectively. The general rule is the following: when we have a case of independent misclassification and the ranks of the misclassification matrices are the number of categories of the

original variables, for those models of the original table with the true variables appearing at most once in the parenthesis, we can substitute these variables by their false variables and get the equivalent models. Therefore, for these equivalent models, we can test them without knowing the misclassification matrices.

Thus, knowing the values in the misclassification matrices does not help us in discriminating among these models. However, it enables us to estimate measures of association and the original cell probabilities.

In a three-dimensional $J \times K \times L$ table, if the variable j is randomly misclassified into a variable i , then we have an overall model $H_{(12,234)}$. From the rule given above, we know that $H^*_{(23,24)}$ is not equivalent to $H^*_{(13,14)}$. This is the categorical analog of a statement given in the section 32 of Fisher (1970): an apparent partial correlation between the two variates will, therefore, be produced by random errors in the third variable.

5. Various Randomized Response Schemes.

In this section, we want to consider some other randomized response schemes and their associated misclassification matrices.

For the unrelated question RR scheme mentioned in the Introduction section, if the population proportion of "yes" to the unrelated question (π_y) is unknown, then we have to separate the sample randomly into two groups ($K=2$) and the first (or second) group will select the original sensitive question to answer with a probability p_1 (or p_2) and the unrelated question to answer with a probability $1-p_1$ (or $1-p_2$). Let the true answer of the unrelated question be $l (L=2)$, then the misclassification matrix from (j, k, l) to i is

	j=1				j=2			
	k=1		k=2		k=1		k=2	
	l=1	l=2	l=1	l=2	l=1	l=2	l=1	l=2
i=1	1	p_1	1	p_2	$1-p_1$	0	$1-p_2$	0
i=2	0	$1-p_1$	0	$1-p_2$	p_1	1	p_2	1

Assuming there is not other variable of interest, the overall model will be

$$\pi_{ijkl} = \pi_{+jkl} \left(\pi_{ijkl} / \pi_{+jkl} \right) \quad (5.2)$$

The observed data $\{c_{ik}\}$ is again a multinomial with cell probabilities $\{\pi_{i+k+}\}$. In this case, in order to estimate $\{\pi_{+jkl}\}$ uniquely, we have

to begin with the model $H^*_{(2,3,4)}$ which means the answer to the sensitive question, the random subgrouping, and the answer to the unrelated question are mutually independent. From this RR scheme, we are not able to differentiate between models $H^*_{(3,24)}$ and $H^*_{(2,3,4)}$.

Folsom *et al.* (1973) propose using two alternate unrelated questions in addition to the sensitive question and setting up a randomized scheme as follows. For a dichotomous stigmatizing trait (k), we will take two samples of the population and use two other dichotomous unrelated questions (l, m) with unknown distributions. For the first sample, the subject will answer the question m truthfully and, in addition, answer the sensitive question k with a probability p and answer the question l with a probability $1-p$. Let us call the observed randomized response as i . For the second sample, the subject will answer in the same way except with questions l and m interchanged. Let us call the observed randomized response as j . Let us use the superscript within parenthesis to indicate two samples, then the overall model will be

$$\pi_{i+klm}^{(1)} = \pi_{++klm} \left(\pi_{i+lk+} / \pi_{++kl+} \right), \quad (5.3)$$

$$\pi_{+jklm}^{(2)} = \pi_{++klm} \left(\pi_{+jk+m} / \pi_{++k+m} \right),$$

$$\text{for all } i, j, k, l, m. \quad (5.4)$$

The misclassification matrix for the first sample is

	k=1		k=2	
	l=1	l=2	l=1	l=2
	i=1	1	p	$1-p$
i=2	0	$1-p$	p	1

$$(5.5)$$

and the misclassification matrix for the second sample will be the same except that l will be replaced by m . Let the observed counts be $\{c_{im}^{(1)}\}$ and $\{c_{jl}^{(2)}\}$ respectively, and the overall total for two samples be N . The ML equations under the model $H^*_{(345)}$ are

$$\hat{\pi}_{++klm} = \sum_i \left[c_{im}^{(1)} / (N \hat{\pi}_{i+++m}) \right] \hat{\pi}_{i+klm}^{(1)} + \sum_j \left[c_{jl}^{(2)} / (N \hat{\pi}_{+j+l+}) \right] \hat{\pi}_{+jklm}^{(2)},$$

$$\text{for all } k, l, m, \quad (5.6)$$

and $\{\hat{\pi}_{i+klm}^{(1)}\}$ and $\{\hat{\pi}_{+jklm}^{(2)}\}$ should satisfy

equations (5.3) and (5.4), i.e.,

$$\hat{\pi}_{i+k|lm}^{(1)} = \hat{\pi}_{++k|lm} \left(\frac{\pi_{i+k|l+}}{\pi_{++k|l+}} \right), \quad (5.7)$$

and

$$\hat{\pi}_{+jk|lm}^{(2)} = \hat{\pi}_{++k|lm} \left(\frac{\pi_{+jk|m}}{\pi_{++k|m}} \right), \quad (5.8)$$

A solution can be obtained iteratively. However, we will have identifiability problem unless we consider some other simpler models than H^* (345).

Lin and Chow (1976b) consider the multiple trials RR scheme. They apply the same RR scheme to a variable k twice to obtain two false variables i and j . The overall model is

$$\pi_{ijk} = \pi_{++k} \left(\frac{\pi_{i+k}}{\pi_{++k}} \right) \left(\frac{\pi_{+jk}}{\pi_{++k}} \right). \quad (5.9)$$

The observed data $\{c_{ij}\}$ is a multinomial with cell probabilities $\{\pi_{ij+}\}$. The last two terms in equation (5.9) correspond to an identical misclassification matrix.

For other schemes (Abul-ela, *et al.*, 1967; Warner, 1971; Hochberg, 1975; Lin, Chow, and Mosley, 1975; Lin and Chow, 1976a), the misclassification matrix can be obtained accordingly. If we limit the discussion to one true variable, then our iterative procedure will produce the maximum likelihood estimates of proportions. For some rare data, the estimates given in various papers will produce inadmissible value as Singh (1976) has pointed it out. Our iterative values will converge to the boundary of the parameter space with the estimates of some cell probabilities approach to 0.

Another approach to fit log-linear models for the randomized response data, by viewing the data geometrically as a mixed-up table and not considering the misclassification matrix explicitly, is given in Chen (1978).

6. Conclusion. This paper develops a flexible and usable methodology for the analysis of contingency tables derived from randomized response. The critical point is that we have misclassification matrices with known numerical values. This methodology will have applications also in epidemiological studies of disease etiology where the classification of disease has known sensitivity and specificity. It can also be applied to missing data problem with known missing probabilities.

The problems which deserve future research are the estimation of variances of the estimates of the various parameters

and probabilities and the investigation of optimal designs for discrimination among models. Lee (1977) has investigated the first problem in the ordinary contingency table without involving misclassification. Leysieffer and Warner (1976) and Loynes (1976) have investigated the second problem about optimal design for the estimation of proportions.

REFERENCES

- Abul-ela, A.A., Greenberg, B.G., and Horvitz, D.G. (1967), "A Multi-proportions Randomized Response Model," *J. Amer. Statist. Asso.*, 62, 990-1000.
- Assakul, K., and Proctor, C.H. (1967), "Testing Independence in Two-Way Contingency Tables with Data Subject to Misclassification," *Psychometrika*, 32, 67-76.
- Bishop, Y.M.M., Fienberg, S.E., and Holland, P.W. (1975), *Discrete Multivariate Analysis: Theory and Practice*, MIT Press, Cambridge, Massachusetts.
- Bross, I. (1954). "Misclassification in 2 x 2 Tables," *Biometrics*, 10, 478-486.
- Chen, T.T. (1978), "Log-linear Models for the Categorical Data Obtained From Randomized Response Techniques," *1978 Social Statistics Section, Proc. Amer. Statist. Asso.*, 284-288.
- _____ (1979), "Log-linear Models for Categorical Data with Misclassification and Double Sampling," *J. Amer. Statist. Asso.*, 74, 481-488.
- _____ and Fienberg, S.E. (1976), "The Analysis of Contingency Tables with Incompletely Classified Data," *Biometrics*, 32, 133-144.
- _____ and Veale, J.R. (1974), "Preliminary Results of the Survey of CSUH Students, Technical Report, Dept. of Statistics, California State University, Hayward.
- Cochran, W.G. (1977), *Sampling Techniques*, 3rd. ed., John Wiley, New York.
- Dalenius, Tore (1977), "Privacy Transformations for Statistical Information Systems," *J. of Statist. Planning and Inference*, 1, 73-86.
- Drane, W. (1976), "On the Theory of Randomized Responses to Two Sensitive Questions," *Communications in Statistics*, A5(6), 565-574.
- Fienberg, S.E. (1977), *The Analysis of Cross-classified Categorical Data*, MIT Press, Cambridge, Massachusetts.

- Fisher, R.A. (1970), Statistical Methods for Research Workers, 14th ed., Hafner Press.
- Folsom, R.E., Greenberg, B.G., Horvitz, D.G., and Abernathy, J.R. (1973), "The Two Alternate Questions Randomized Response Model for Human Surveys," J. Amer. Statist. Asso., 68, 525-530.
- Goodman, L.A. (1973), "The Analysis of Multidimensional Contingency Tables When Some Variables Are Posterior To Others: A Modified Path Analysis Approach," Biometrika, 60, 179-192.
- ____ (1974a), "The Analysis of Systems of Qualitative Variables When Some Of the Variables Are Unobservable: Part I - A Modified Latent Structure Approach," Amer. J. Socio., 79, 1179-1259.
- ____ (1974b), "Exploratory Latent Structure Analysis Using Both Identifiable and Unidentifiable Models," Biometrika, 61, 215-231.
- Greenberg, B.G., Abul-ela, A.A., Simmons, W.R., and Horvitz, D.G. (1969), "The Unrelated Question Randomized Response Model: Theoretical Framework," J. Amer. Statist. Asso., 64, 520-539.
- Haberman, S.J. (1974a), The Analysis of Frequency Data, University of Chicago Press, Chicago.
- ____ (1974b), "Log-linear Models For Frequency Tables Derived By Indirect Observation: Maximum Likelihood Equations," Annals of Statist., 2, 911-924.
- ____ (1975), "Iterative Scaling Procedure For Log-linear Models for Frequency Tables Derived by Indirect Observation," 1975 Statist. Computing Sec. Proc. Amer. Statist. Asso., 45-50.
- Hochberg, Y. (1975), "Two Stage Randomized Response Schemes for Estimating a Multinomial," Communication in Statistics, 4(11), 1021-1032.
- Lee, S.K. (1977), "On the Asymptotic Variances of \hat{u} Terms in Log-linear Models of Multidimensional Contingency Tables," J. Amer. Statist. Asso., 72, 412-419.
- Leysieffer, F.W. and Warner, S.L. (1976), "Respondent Jeopardy and Optimal Designs in Randomized Response Models," J. Amer. Statist. Asso., 71, 649-656.
- Liu, P.T., Chow, L.P. and Mosley, W.H. (1975), "Use of the Randomized Response Techniques with a New Randomizing Device," J. Amer. Statist. Asso., 70, 329-332.
- ____ and Chow, L.P. (1976a), "A New Discrete Quantitative Randomized Response Model," J. Amer. Statist. Asso., 71, 72-73.
- ____ and Chow, L.P. (1976b), "The Efficiency of the Multiple Trial Randomized Response Technique," Biometrics, 32, 607-618.
- Loynes, R.M. (1976), "Asymptotically Optimal Randomized Response Procedures," J. Amer. Statist. Asso., 71, 924-928.
- Mote, V.L., and Anderson, R.L. (1962), "An Investigation of the Effect of Misclassification on the Properties of χ^2 Test in the Analysis of Categorical Data," Biometrika, 52, 95-109.
- Rubin, T., Rosenbaum, J., and Cobb, S. (1956), "The Use of Interview Data for the Detection of Associations in Field Studies," Journal of Chronic Diseases, 4, 253-266.
- Singh, J. (1976), "A Note on Randomized Response Techniques," 1976 Social Statist. Sec. Proc. Amer. Statist. Asso., 772.
- Warner, S.L. (1965), "Randomized Response: A Survey Technique for Eliminating Evasive Answer Bias," J. Amer. Statist. Asso., 60, 63-69.
- ____ (1971), "The Linear Randomized Response Model," J. Amer. Statist. Asso., 66, 884-888.