SURVEY OF INCOME AND PROGRAM PARTICIPATION

Working Paper Series

Response and Procedural Error Variance in Surveys: An Application of Poisson and Neyman Type A Regression

No 9016 121

Daniel H. Hill
Opinion Research Institute
University of Toledo

December 1990

The author is Director of the Opinion Research Institute and Associate Professor of Economics at the University of Toledo, and Associate Research Scientist at the Survey Research Center of the University of Michigan. This research was sponsored in part by Joint Statistical Agreements (JSA-87-5) between the United States Bureau of the Census and the Survey Research Center of the University of Michigan and between the United States Bureau of the Census and the Opinion Research Institute of the University of Toledo (JSA-90-01). The author would like to thank Dan Kasprzyk, Fred Cavanaugh, Chet Bowie, Jeff Moore, Irv Schreiner, Gary Shapiro and Vicki Stout of the Census Bureau, Jim Lepkowski and Graham Kalton of the Survey Research Center, Jim LeSage of the Department of Economics of the University of Toledo, and Clifford Clogg of the Department of Statistics of Pennsylvania State University for their helpful comments on earlier versions of this paper. Any errors are the responsibility of the author.

TABLE OF CONTENTS

Response and Procedural Error Variance in Surveys: An Application of Poisson and Neyman Type A Regression

Daniel H. Hill

1. Introduction

Most studies of non-sampling errors in surveys employ interview/reinterview data to estimate and analyze response variances. Because the respondent is most often the focus of attention, these studies restrict the samples to those cases where, what O'Muircheartaigh (1986) terms, 'the essential survey conditions' are the same for both the interview and reinterview observations. The variances analyzed are, therefore, conditional on these essential survey conditions. One such condition is that the respondent actually is asked the question in both trials. In fact, however, a potentially important source of unreliability of survey data is that the procedures used in determining who gets asked which questions are themselves subject to error. Thus, the unconditional variances are also relevant in assessing the quality of survey data.

In this paper we assess the relative importance of response and procedural error variance in the Survey of Income and Program Participation (SIPP). We concentrate on the SIPP not because it is any more prone to error than other surveys, but rather because it has a good reinterview program and the staff has had the curiosity to encourage external scrutiny of these data. We conduct our assessment in two ways. First, (Section 3) we examine each of a series of questionnaire items from the SIPP Reinterview Program and present separate estimates for the response variance, the procedural variance, and the overall variance. Second, we investigate the correlates of response and overall error variance by modeling the entire interview/reinterview outcome as single experiment. A very general counting distribution, the Poisson-Pascal, is taken as a point of departure (Section 4) in analyzing the number of interview/reinterview discrepancies. This distribution subsumes a large family of more specific distributions (see Katti and Gurland, 1961) some which allow for heterogeneity and/or contagion (*e.g.* the Negative-Binomial or Pascal and the Neyman Type A) and some which do not (*e.g.* the Poisson). The object of this preliminary analysis is to identify the most parsimonious distribution which adequately describes the data. Once this is accomplished, the model's parameters are treated as functions of the characteristics of the respondent and the interview situation and the effects of these characteristics on data quality are estimated.

2 Background

Before beginning our analysis a certain amount of background may be useful. The prevalence and nature of non-sampling errors is of considerable concern to both the producers and consumers of survey data. As producers, to improve the quality of data we need to know what types of questions are most prone to error and for what types of respondents. We also need to know the extent to which the errors in our data are a consequence of our own procedures or are a reflection of inherent imperfections in the response process. As consumers of survey data we need to know many of the same things about these measurement errors so that we can more realistically interpret our findings and adjust our estimating procedures to ameliorate the effects of the errors on our estimates.

While record check studies are potentially the best source of estimates of survey errors, they are quite expensive and often can be performed for only special populations (see, *e.g.* Duncan and Hill, 1986, 1989). A far more common source of data allowing the estimation of response variance is the interview/reinterview method (see *e.g.* Bailar, 1968, O'Muircheartaigh, 1976, 1984). Many large-scale surveys conduct reinterviews of subsamples of respondents as a part of their quality control programs and these are often of sufficient quality to allow estimation of response variance.

2.1 A Model of Response Error

The underlying response model in most response error studies is similar to that first proposed by Hansen, Hurwitz, and Bershad (1961). According to this model the value recorded for respondent j for question i at time t is

$$y_{ijt} = y_{ij} + \beta_{ij} + \epsilon_{ijt}$$

where y_{ij} is the true value, β_{ij} is a fixed response error or bias, and ϵ_{ijt} is the variable response error. With observations at two points in time it is possible to use the paired difference $y_{ij1} - y_{ij2}$ for $\epsilon_{ij1} - \epsilon_{ij2}$. Assuming $\sigma^2_{\epsilon_{i1}} = \sigma^2_{\epsilon_{i2}} \equiv \sigma^2_{\epsilon_i}$ then the variance of the paired difference is:

$$\sigma^2_{\epsilon_{i1,2}} = 2\sigma^2_{\epsilon_i}(1 - \rho_{i1,2})$$

where $\rho_{i1,2}$ is the between trial correlation in variable errors for question i due to respondents' memory of the interview responses. While Hansen, Hurwitz and Bershad assumed this correlation to be zero, more recent work has attempted to estimate it (see e.g. O'Muircheartaigh, 1986). Thus, a conservative estimate of the simple response variance (SRV) is simply one-half the variance of the paired difference. With discrete data this corresponds to one-half the gross-difference-rate (GDR *i.e.* the number of discrepant paired observations divided by the total number of paired observations).

2.2 The SIPP Reinterview Program

The SIPP reinterview program is an ongoing systematic operation which is intended to monitor data quality by checking the interviewers' work. The sample to be reinterviewed each month is a multistage probability sample of current SIPP respondents.

Figure 1 illustrates the question flow for Section 2 of the Reinterview Questionnaire. The questions actually asked of the respondent in both the interview and reinterview are printed in bold, while the Office Check Items which are transcribed to the Reinterview Questionnaire from the original appear in normal print. Unless otherwise indicated, questions are asked in sequence. In most cases, however, respondents are skipped around certain questions and these skips are indicated in the figure by lines and arrows. If, in response to question 1, for instance, the respondent said he had a job for at least part of the reference period ('yes' on item 1.), he is skipped around the questions about whether he spent any time looking for a job (2a.), or whether he wanted a job (3a.), and is asked about whether he had a job each week of the reference period instead (4.). In Figure 1, a skip such as this which results from a response to a question asked in the reinterview study is depicted with a dotted line. Skips from Office Check Items, being automatic from the reinterviewer's point of view, are depicted as solid lines.

[Figure 1 about here]

It does not take a great deal of study of Figure 1 to see that the skip sequences employed in the SIPP can be quite complicated. Indeed, a major goal of the reinterview program is to see if individual interviewers are following these skip sequences properly.

3. Inconsistency Rates and Simple Response Variance Estimates

With the two observations provided by the interview and reinterview responses it is possible to estimate the simple response variance for the various questions. To do so, we first confine our attention to that portion of the reinterview sample where a) the reinterview was successfully conducted and b) it was determined that the interviewer had visited the proper sample unit in conducting the original interview. These restrictions leave us with a sample of 1,559 cases of interview/reinterview data for waves 2 and 3 of the 1984 panel.

As noted in the introduction we can distinguish two distinct types of errors when the interview and reinterview reports do not agree—response errors and procedural or 'skip' errors. We will reserve the term response variance or 'response errors' for estimates involving cases where the question was actually asked of a respondent in both the interview and reinterview and where a response was recorded. Given the complicated skip sequences employed, it should not be surprising that there are differences between the two reports not just in responses, but in whether or not the question was asked each time. Discrepancies between the interview and reinterview arising because a question was skipped in one and not the other will be referred to as 'procedural errors'.

Neither of the terms 'response' or 'procedural' in referring to errors should be taken too literally, however. Response errors can come about, for instance, because the interviewer marked the wrong answer (a procedural error), and procedural errors can appear because of a error in an earlier answer provided by a respondent.

An example may be useful in clarifying these distinctions. Table 1 presents the recorded responses for the interview and reinterview for Item 4.—the question regarding receipt of state unemployment compensation. Actual responses in both interviews were recorded for only some thirteen percent (=100*207/1559) of the cases. Of these 2.9% (=100*(3+3)/207) of the reports were different. The simple response variance for this question is, therefore, .0145, or half the gross difference rate among those respondents who answered the question in both the interview and reinterview. We will define the procedural error rate as the simple gross difference rate for whether the question was asked. For the unemployment compensation question results in Table 1, the procedural error rate is 6.54 percent (= 100*(7+59+7+29)/ 1559). The overall error rate is simply the fraction of the entire sample for which the interview and reinterview reports differ. It is equal to the sum of the procedural error rate and the response error rate, with the latter weighted by the fraction of the sample with valid responses in both interviews. That is, for each question j:

$$OER_j = PER_j + RER_j * DR_j$$

where OER is the overall error rate, PER is the procedural error rate, RER is the response error rate, and DR is the dual response rate.

[Table 1 about here]

Table 2 presents these error rates and the dual response rates for each of twelve substantive questions asked in the SIPP reinterview. There is considerable variation in the overall error rates for these questions ranging from less than two percent for questions on employment during the reference period (1) and continued Medicaid coverage (26b) to about seven percent for the Health Insurance coverage (27a) and the employer's contribution to Health Insurance (27f) questions. This pattern is quite similar to that reported by the Census

Bureau's Reinterview Evaluation Section (see e.g. Smith, 1987). The overall error rate over all items was 3.82% which is only moderately higher than the 3.07% reported by St. Clair (1985) for Waves 2–4 of the 1984 Panel. Most of this difference is probably due to differences in the definitions of difference rates. It is also likely, given the results of Section 4 below, that our rate would have been lower had we included wave 4 in our analysis.

While it does vary from question to question, the majority of the errors in the data as a whole are procedural rather than response errors. Given the skip patterns depicted in Figure 1, it is not surprising that virtually all of the errors on the Medicare coverage question were procedural in nature—i.e. the result of the question being skipped in one interview and not in the other. There are, after all, three distinct ways in which a respondent can be routed around question 23a and four ways in which he could be routed to it.

[Table 2 about here]

Procedural errors also accounted for most of the overall errors in all the remaining questions except for the initial employment and health insurance questions. That these are the initial questions in a sequence which all respondents are to be asked is significant and points to the fact that some of the procedural errors are the result of response errors in earlier portions of the interview.

Response errors also vary widely from a low of less that three-tenths of one-percent for the Foodstamp authorization question to more than seven and a half percent for the employer health insurance contribution question. The high response variances of health insurance coverage and employer contribution of .03 (=.5*6.03/100) and .038 (=.5*7.62/100), respectively, would suggest that there is something wrong with these questions. The full health insurance coverage question reads:

> 27a)**"During the 4-month period, did ... have group or individual health insurance in ...'s own name?"**

While the problem with this question is quite likely that 'whose name the insurance is in' is not particularly salient or important to the respondent, it would be interesting to know how many respondents gave either "group" or "individual" as their initial response. Similarly, from the respondent's point of view, reasonable responses to the question:

> 27f)**"Did the employer or union (former employer or pension plan) pay for all or part of the cost of this plan?"**

could be 'employer', 'union', 'all', 'part', 'no', or 'yes'. The allowed responses are 'all', 'part' and 'none'. Thus, it is quite likely that the interviewer is having to probe for the 'all', 'part', or 'none' responses in a large number of cases when the respondent's answer is 'yes', 'employer' or 'union'. Part of the response variance may be due to variance in how and whether these probes are being made.

While response variance is most troublesome for the health insurance questions, it is also quite high for discouraged worker question. In this case, the question seems rather unambiguously worded and it would seem that the problem must lie in the ambiguity of the concept itself.

Before leaving our discussion of the extent of interview/reinterview errors it should be noted that independent analyses of the reinterview data by Bureau staff revealed the same pattern of results for the health and discouraged worker questions. As a result the health questions have been substantially modified, while the discouraged worker question has been dropped.

In summary, simple comparisons of interview and reinterview reports from the reinterview data are sufficient to highlight some questions and procedures that are particularly problematic in the current SIPP instrument. Considerable error is probably being introduced to the data, for instance, because the skip sequences are sometimes quite complex and may not always be successfully followed. Additional errors occur because not all the questions are as clearly worded as we would like, and the reinterview data reflect these glitches in the form of high response variance.

4. Correlates of Inconsistency

If the procedural and response variability is the same for all respondents, then its existence is relatively benign. In multivariate analysis its existence in dependent variables will only reduce the model's goodness of fit and in independent variables will (predictably) bias the estimated coefficients toward zero. If, on the other hand, the extent of response or procedural variance differs systematically from one respondent to the next, all manner of problems can be expected to arise in bivariate or multivariate analysis. The purpose of this section is to explore the extent to which response and procedural variance differs systematically with characteristics of respondents and interviewers.

Traditionally, analysts have chosen some form of logit model (see e.g. O'Muircheartaigh and Wiggins, 1981) in investigating the association of respondent and interviewer characteristics with response errors. Such analyses are done on a question-by-question basis.

In a preliminary investigation of such a model with the current data, the author found that, given the rarity of response errors and the relatively small size of the SIPP reinterview program, there were too few cases of response errors to analyze effectively in this manner.

An alternative modeling approach is to analyze the reinterview data, not on a question-by-question basis, but as single experiment in which the outcome is the number of errors occurring in the course of the reinterview. This total count of errors can be thought of as the outcome of a discrete stochastic process and should be describable in terms of a discrete distribution. Determining which of the infinite number of such distributions is most appropriate for the response and total error data is a difficult task. Fortunately, there are some very general forms which subsume many more specific distributions as limiting cases. By estimating the parameters of the former we can identify which of the latter to use in our analysis of the correlates of response and total error variances.

The general distribution we will investigate is the Poisson-Pascal which, as Katti and Gurland (1961) demonstrate, has among others the Negative-Binomial, Neyman A, Poisson and Pòla-Aeppli distributions as limiting cases. The differences between these distributions is a result, primarily, of differences in the extent and nature of the dependence between the probabilities of the various events. In our case, if an error in one question has no effect on the probability of an error in another, and if there are no other 'order effects', then the stochastic process producing the count data can be thought of as a Poisson process. If, on the otherhand, errors are associated then a more complicated stochastic process is involved. The Poisson-Pascal distribution is consistent with errors occurring in clusters or 'clumps' (in the case of heterogeneity) or in 'out-breaks' (in the case of contagion). The probability of encountering such clumps or outbreaks of errors is distributed Poisson while the number of erroneous cases within an outbreak is distributed according to the negative-binomial or Pascal distribution.

The Poisson-Pascal distribution is therefore a compound distribution involving Poisson mixing of the Pascal. The probability generating function is:

$$\exp[\lambda\{((1+\delta) - \delta t)^{\theta} - 1\}]$$

.1)

where λ is the Poisson parameter and θ and δ are the parameters of the Pascal. From this one can derive the following recursions:

$$P(x=0) \equiv P_0 = \exp[-\lambda(1-(1+\delta)^{-\theta})] \quad 4.2)$$

and

$$P(x=k) = P_k = [\exp(-\lambda)(\delta/(1+\delta))^k/k!]\,[\sum_{j=1}^{\infty}(\theta j)^{[k]}\{(\lambda(1+\delta)^{-k})^j/j!\}] \quad 4.3)$$

where $Z^{[j]}$ represents the ascending factorial operator $Z*(Z+1)*....*(Z+j-1)$.

Katti and Gurland (1961) show that as $\theta \longrightarrow \infty$ and $\delta \longrightarrow 0$ the Poisson-Pascal approaches the Neyman A. Similarly as $\delta \longrightarrow 0$ and $\lambda \longrightarrow \infty$ the distribution approaches the Poisson. Finally the negative-binomial or Pascal results when $\theta \longrightarrow 0$ and $\lambda \longrightarrow \infty$.

While statisticians have developed some rather sophisticated techniques for discriminating between various discrete distributions (see *e.g.* Johnson and Kotz, 1969 page 42 – 46), they are not as amenable to formal testing as is straight forward application of maximum likelihood techniques to the Poisson-Pascal and its limiting forms. In this case we can perform exact likelihood-ratio tests of the limiting-case restrictions. The likelihood-ratio of interest reduces to:

$$2*(\ln(L^{PP}) - \ln(L^d)) \sim \chi^2 \text{ d.f.} = 3 - \text{d.f.(d)} \quad 4.4)$$

where $\ln(L^k)$ is the log-likelihood of distribution k obtained by maximizing:

$$\ln(L^k) = \sum_{i=1}^{N} \ln(P^k_{n_i}) \quad 4.5)$$

with respect to the d.f.(k) parameters of distribution k. Appendix A describes the maximum-likelihood algorithm used for the Poisson-Pascal and compares its results to those obtained Katti and Gurland's (1961) method of moments. The maximum-likelihood technique results in a better fit to the data.

[Figure 2 about here]

Table 3 presents maximum-likelihood estimates of the parameters of the Poisson-Pascal, and several of its limiting forms, for response and total error counts in the SIPP data. Interestingly, as Figure 2 verifies, all the distributions investigated fit the response-error count data equally well. Evidently, the stochastic process generating these data is a particularly simple one—the Poisson—which is a limiting form of each of the more complex processes. Indeed, the response-error data are so well described by the Poisson that it becomes difficult to obtain convergence of the log-likelihood function for the more general distributions. In essence the maximization algorithm neither needs nor can well use more than one parameter to fit these data. Nevertheless, convergence was eventually obtained for each of the distributions listed in Table 3.

[Table 3 about here]

The fact that in the Poisson-Pascal case δ is small and λ large suggests the Poisson for the response error data, as does the fact that the mean and variance are equal. When we estimate the Neyman A for response error counts we find θ' (the index of 'clumping' or 'contagion') to be so small as to be negligible. Similarly, the shape parameter in the ordinary Pascal δ'' is very large (140.8) which implies negligible heterogeneity of the response errors. The likelihood-ratio tests values are essentially zero with 1, 1, and 2 degrees of freedom for the Neyman A, Pascal and Poisson, respectively. Thus, we can reject none of the restrictions of the limiting distributions and will use the Poisson (*i.e.* the most parsimonious) below in modeling response errors.

A Poisson process is a maximal disorder process and the fact that it fits the response error data so well suggests that true response errors occur infrequently and at random. Such a simple process, however, is not adequate to explain total errors. There are two reasons for this; as a re-examination of Figure 1 will suggest. First, the skip sequences vary in complexity from one part of the reinterview to the next. This suggest that when we are considering both response and procedural errors there will be significant heterogeneity. Second, and perhaps more importantly, errors at one point in the reinterview may be a direct result of errors at a preceding point. This is contagion. Both the Poisson-Pascal and the Neyman Type A process allow for such a complex combination of heterogeneity and contagion, while neither the ordinary Pascal not the simple Poisson do (see Figure 3). The technical difficulties encountered in maximizing the likelihood function for the response error data do not appear in the case of the total-error data.

[Figure 3 about here]

The likelihood-ratio test value for the Neyman A restriction is virtually zero (0.36) with one degree of freedom, and, hence we can not reject it. We can, however, soundly reject the Pascal and Poisson restrictions for the total error data since the likelihood-ratio test value of the former is approximately 22 with 1 degree of freedom and in excess of 780 (d.f. = 2) for the latter.

[Figure 4 about here]

As in the case of response errors, we will use the most parsimonious distribution which is consistent with the total error data in our subsequent analysis. This is the Neyman A. This stochastic process suggests that errors occur in clusters or 'outbreaks' with the number of clusters being distributed Poisson (with $\lambda = .54$) and the number of errors within an clusters also being Poisson (with $\theta = 1.06$). Figure 4 depicts these two Poisson distributions which together describe the total-error process. The distribution of clusters is quite leptokurtic with its mode at zero—implying that the majority of cases have no errors at all. The distribution of errors within clusters, by contrast, is more platokurtic with its mode at one. Nearly ten-percent of the outbreaks have three or more errors within them.

4.1 Application of Poisson Regression to Response Errors

Conceptually, the nearly perfect fit of the response inconsistency data to the Poisson suggests that if respondents were asked a reinterview question repeatedly (and their memories of their previous responses were wiped clean) inconsistent reports would appear infrequently, randomly and independently in time. Indeed, the Poisson can be shown to be the maximum entropy or disorder process. Furthermore, if errors in each question are the result of a Poisson process then the sum of the response errors throughout the reinterview will also be Poisson. With a Poisson process, the probability of exactly n_i errors occurring is:

$$p(n_i) = \exp(-\lambda)\, \lambda^{n_i}/n_i! \qquad 4.6)$$

We can investigate the correlates of response errors in the context of a Poisson process by simply allowing the Poisson parameter (λ_i) to vary with characteristics (X_i) of the individual case according to:

$$\lambda_i = Q_i \exp(X_i \Lambda) \qquad 4.7)$$

where Λ is a parameter vector reflecting the effects of characteristics on mean response error rates and Q_i is the number of questions actually asked of respondent i. We use Q_i as a weighting factor in order to purge the estimates of Λ of any biases due to the fact that not all respondents are asked the same number of questions. This causes differing exposures to the risk of errors which may be correlated with the X_i.

Expressions 4.6) and 4.7), in conjunction with 4.5), form the basis of what is sometimes referred to as Poisson Regression. There are several attractive features of Poisson regression in analyzing response errors. First, Poisson Regression is a well established tool which has successfully been employed in a wide range of fields (see *e.g.* Maddala, 1983, Hausman, Hall and Griliches, 1984, and Haight, 1967). Indeed, the Poisson error structure in conjunction with the log-link of equation 4.2 forms the basis of the entire class of log-linear models (see *e.g.* McCullagh and Nelder, 1989). This also means that the model can be estimated with a number of standard packaged programs such as GLIM and LIMDEP. Additionally, in the context of the present problem, the effects of independent variables are easily interpretable and can be readily compared with the results of other analyses in the literature. To see this, note that:

$$\lambda_i = \sum_{q=1}^{Q_i} GDR_q/Q_i = 2\sum_{q=1}^{Q_i} SRV_q/Q_i$$

where GDR_q and SRV_q are, respectively, the gross-difference rate and simple response variances for the q^{th} question.

The independent variables we employ in our analysis are of two types—those intended to capture (at least some of) the effects variability in interviewing process, and those characteristics of respondents which might affect response variability. The first of the interviewing process variables is simply the calendar month in which the original interview was taken. Since the data are taken from the second and third waves of the 1984 panel, the study was still quite new to the interviewers at the beginning of our observation period. We would expect more errors in these months. By the end of our observation period, on the other hand, most interviewers had been administering the study monthly for a full year, and we would expect their error rates to have settled down. Because we would expect declining marginal improvements with additional months of experience, we include the natural logarithm of the interview month rather than the month itself in our empirical specification.

The second interviewing process variable is a scale based on the overall performance of interviewers in the various Regional Offices. The underlying rationale for this scale is that an unknown portion of the observed variation between these offices is due to differences in interviewers and in local procedures and the remainder is due to differences in the characteristics of the respondents. If all of the individual-to-individual variability is due to these Regional Office factors, then a scale constructed from the Regional Office rates should bear a one-to-one relationship with the individual error rates, and should explain all of the variance in them. If on the other hand, the reason Regional Offices differ is that the characteristics of their respondents differ then the one-to-one relationship between the

Regional Office rate and the individual rates should disappear once the individual factors are controlled.

The third and forth interviewing process variables included are the relationship of the individual to the household reference person, and a dummy variable for whether a proxy informant was used in the original interview. The relationship to reference person measure is also a dummy variable equaling 1 if the individual is some one other than the reference person or his/her spouse (*e.g.* child, parent, aunt, *etc.*).

The individual characteristics included in our empirical specification are the same ones thought to affect market productivity in the human-capital model of earnings. These consist of age (and its square), education, race, and gender. We also include income itself in some of our specifications.

[Table 4 about here]

Table 4 presents both bivariate and multivariate estimates of the Poisson Regression model for response errors obtained by maximizing 4.6) with respect to the Λ. The first column of figures, labeled 'Bivariate Parameters', are obtained when the Poisson Regression model is estimated with only a constant and the variable listed to the left of the coefficient included as predictors. As hypothesized, response errors decline significantly with interview month. Since the month is included as a proxy for interviewer and respondent experience with the SIPP, and since the logarithm of month is used, the coefficient of $-.275$ is interpretable as the experience elasticity of response error—a one percent increase in experience is associated with a .275 <u>decrease</u> in response error rates. This result is encouraging because it indicates that progress was being made in improving response quality early in the SIPP program.

The fact that the coefficient on the log on the Regional Office error rate is so close to unity, and is highly significant means that differences in something at the regional level are important, but the bivariate results can provide no clue as to what it might be. While the effect of the original interview having been taken with a proxy respondent is to increase response error, the effect is not sufficiently strong to attain statistical significance. The positive coefficient for the relationship to reference person dummy variable indicates that the response consistency for reference persons and their spouses is higher (by about 38.3 percent) than that obtained from other persons in the household.

The effects of age on response error rates is highly non-linear. The coefficients of $-.485$ and .047 on age and age square, respectively, suggest that response quality increases with age at a decreasing rate until age 51 where it attains its maximum. For respondents much older or younger than this, response quality is significantly lower. While in the present case it is clear from the individual coefficient's standard errors that the age effects are significant, in general, one would need to test the change in the goodness of fit when age and its square are dropped out of the analysis as a set. This can be accomplished by means of a likelihood-ratio test constructed from the log-likelihood values presented in the second column of figures. In the present case the χ^2 associated with the hypothesis that age (and its square) are not associated with response quality is 8 (= 2*(−771.4 − (−775.4))), and has 2 degrees of freedom. Thus, the null hypothesis of no age effect can be rejected soundly.

The final two variables with significant bivariate associations with response errors are education and income. Each one-year increase in educational attainment is associated with a 4.4 percent decrease in the response error rate. The extremely significant coefficient of −.827 on income, similarly, is interpreted as indicating that a dollar increase in monthly personal income is associated with a .83 percent decrease in the response error rate. Monthly personal income is the most powerful predictor of response errors included in our analyses. Conceivably some of this effect may be a reflection a tendency for fewer imputations being made for relatively complete interviewers and these interviewers tend to be interviews with people who have some income to report.

The bivariate results just discussed are analogous to simple correlations in linear models. The multivariate results presented in the last two columns of Table 3, in contrast, are analogous to multiple correlation coefficients. These coefficients are, therefore, interpretable as the net effects of the various factors on response error one obtains when the effects of other factors are controlled. Thus, it is not surprising that these multivariate effects are, in general, weaker than their bivariate counterparts. Indeed, with the single exception of the Regional Office error index, all the coefficients in column 3—the specification which includes everything but income—are of the same sign as those in column 1, but are smaller in absolute value. The estimated standard errors are also, in general, larger in the multivariate analyses—a second indication that the various predictors are correlated with each other. The decreased size of the estimated effects and their increased estimated variance combine to decrease the significance of almost all predictors in the multivariate analysis which excludes income. The only predictor to go from statistical significance to insignificance, however, is relationship to reference person. This indicates that most of the observed bivariate effect of not being the reference person (or his/her spouse) is, perhaps, due to the fact that most of these other individuals are children and children are younger, less educated and less likely to have income to report than their parents. Once the effects of these correlated factors are controlled, these individuals have response errors which are insignificantly different from those of reference persons (and spouses of reference persons). The combined effect of age and age-squared, by the way, remains significant even though the individual coefficients are not.

When income is added to the multivariate specification of the response error Poisson regression, every other individual characteristic becomes insignificant. Taken literally, this result would suggest that all of the effects of age and education on response quality discussed up to this point are the result of the correlation of these factors with income. We find this result hard to believe. Why income, itself, should have a positive effect on response quality is a mystery. One possibility is that the focus of the SIPP is income and transfer program participation and neither the respondent nor the interviewer may be taking the interview as seriously when the individual has 'nothing to report', than when individual income is substantial.

Before moving on to our analysis of total errors, two further aspects of the multivariate Poisson regression estimates of response errors should be noted. First, the overall goodness of fit of both versions of the multivariate model is highly significant. The χ^2 under the null hypothesis of no association for the model presented in column 3 is 27.8 with 10 degrees of freedom and that for the model in column 4 is 35.2 with 11 degrees of freedom. Second, and of more substantial interest, the coefficient on the Regional Office error index was unaffected by the inclusion of respondent characteristics. In fact, this coefficient increased slightly when the other factors were controlled. This suggests that the source of the regional differences in response errors is something other than regional differences in the characteristics of respondents. One possibility is that the quality of interviewer training or selection varies by region. Alternatively, it may be that the care given to the reinterview program varies from one Regional Office to the next. In either event, future analysis of the reinterview data with data on interviewer characteristics, would seem worthwhile.

4.2 Application of Neyman Type-A Regression to Total Errors

Response errors are relevant when one is trying to understand the response process itself, but in many respects a better measure of the reliability of survey items is the total error rate. This is simply the sum of the procedural rate and the response error rate weighted by the portion of the sample asked the question in both the interview and reinterview. As demonstrated in Section 4.0 above, unlike the response-error rate the Poisson distribution is not a good choice for describing or modeling total errors. The Neyman Type A distribution provides a significantly better fit than either the Poisson or the Pascal distribution. Under this distributional assumption the probability of n_i response and procedural errors occurring can be obtained from the following recursions:

$$P(n_i = 0) = \exp[-\theta'(1 - \exp(\lambda'))] \qquad 4.8)$$

and:

$$P(n_i = k) = (\theta'\lambda'/k)\exp(-\lambda') \sum_{j=0}^{k-1} (\lambda'^j/j!)\, P(k\text{-}j\text{-}1) \qquad 4.9)$$

A regression technique analogous to Poisson Regression, which we will term Neyman A Regression, can be implemented by allowing λ' to vary with X_i according to:

$$\lambda_i' = \exp(X_i\Lambda) \qquad 4.10)$$

In principal, we could also specify θ' as a function of the X_i. In the present application, however, this does not make good sense. This index of clumping, or contagion, is the result of the question sequencing employed in the SIPP and is the same for all respondents.

The log-likelihood function for this Neyman A Regression model is obtained in the normal manner by substituting equation 4.10) into 4.8) and 4.9) and the resulting expressions into 4.5). Maximization of this with respect to the Λ was accomplished using the Goldfarb-Shano version of the Davidson-Fletcher-Powell algorithm provided by Press *et al.* (1987). The results are presented in Table 5.

[Table 5 about here]

The results of the maximum-likelihood Neyman A analysis of total errors (Table 5) look very much like those obtained for response errors using Poisson regression (Table 4). The interpretation of these coefficients is the same as that of the Poisson regression coefficients— for those variables entering linearly (e.g. education), a one unit increase is associated with a proportionate change in the error rate of Λ (a 4.8% decrease for education in the bivariate model). The only real difference between the Poisson regression coefficients for response errors and those of the Neyman Type A Regression for total errors is that the latter have lower estimated variances. The same substantive results hold.

As was the case for response errors, the total error rate declines significantly with time, and there remains a one-to-one relationship between regional office error rates and individual rates. Reference persons (and their spouses) have significantly lower error rates than do more distantly related individuals in the sampling unit, but this is evidently due to their higher income and education and to the fact that they are more apt to be 'middle aged'. error rates decline with age until attaining a minimum at age 45 and increase thereafter. Higher educated individuals have lower total error rates, although this effect disappears if one controls for income (i.e. it is not significant in the multivariate model).

Unlike the Poisson results for response errors, race is a significant correlate of total errors. Blacks have total error rates some twenty-eight percent higher than non-Blacks, and this effect does not appear to be merely a reflection of their lower average educations and incomes. Evidently interviewers are 'hitting the check points' less consistently for Black respondents than they do for non-Black respondents.

Finally, as was the case of response errors, monthly personal income is the strongest predictor of total error rates, and when it is included in the multivariate model along with the other predictors, absorbs most of their effects.

In sum, given the strong similarity of the results of the Poisson regression model of response errors and the Neyman A model of total errors, we are lead to suspect that response and procedural errors share a common causal structure. Whatever this structure is, it evidently involves characteristics of both the respondent and the interviewer (or at least of the Regional Office).

5. Discussion

The results of both Sections 3) and 4) indicate that skip-sequence or procedural errors are a more important source of unreliability in survey items, at least in the SIPP, than are response errors *perse*. That response errors are well described as a Poisson process whereas procedural errors are not means is that, abstracting from skip sequence effects, the occurrence of a response error in one question has no effect on the probability of a response error in a subsequent question. One can easily imagine mechanisms which would result in this not being the case. If a respondent realizes that he made a mistake, for instance, and 'got away with it' on one question, then he might be less careful with subsequent answers. But the close fit of the Poisson to the response error process indicates that there is no net effect of any such mechanisms.

That the inclusion of procedural errors destroys the fit of the Poisson model to the data suggests that the sequencing processes itself acts as a contagious influence on the error probabilities from one question to the next. This raises the possibility that more sequencing is being done in studies like the SIPP than is optimal. This potential problem is analogous to the problem of optimal interviewer workloads when the interviewer acts as a correlating influence for response errors. The trade-off in that case is that per interview training costs decrease with work load while response variance increases. In the present case, the overall interview length can be reduced by skipping entire classes of respondents around questions based on their responses to earlier questions. The resulting interviewing time savings come at a cost of increased error variance and therefore decreased question reliability. As is the case with interviewer workloads, this cost is generally unknown and is often ignored in the survey design process, with the result that sequencing may be over utilized just as work loads are often too high.

Our results, therefore, suggest a couple of lines of future research. First, the importance of procedural errors suggests that it may be worthwhile to conduct micro-simulations of proposed questionnaires before they are implemented in production interviewing. All the questions and skip patterns can be incorporated in a computer program in much the same way a CATI or CAPI instrument is implemented. Then a sample of actual or hypothetical cases can be run through the program with known or hypothetical levels of error variance for each item. The program would yield a total error count for the sample. These total error counts could be used as the dependent variable in a multivariate analysis with the characteristics of the hypothetical respondent being used as independent variables. The result of such an analysis would point to the types of respondents who are particularly likely to have unreliable data. Secondly, the item error variances can be increased sequentially to obtain the marginal importance of each item in determining overall error variances. These marginal effects will point to 'hot-spots' in the questionnaire—*i.e.* questionnaire items that are have powerful effects on overall error rates. These hot-spots, once identified, can be addressed specifically in questionnaire redesign.

A second line of future research suggested by our findings is to determine what it is about the various regional offices which causes them to have such a powerful effect on data quality. Interviewer characteristics, which do vary significantly across the regions, are a prime candidate for study.

References

Bailar, Barbara A., 1968, "Recent Research in Reinterview Procedures", *Journal of the American Statistical Association*, Vol 63, 41–63.

Duncan, Greg J., and Daniel H. Hill, 1986, "An Investigation of the Extent and Consequences of Measurement Errors in Labor-Economic Surveys", *Journal of Labor Economics*, 3, 80–101.

Duncan, Greg J. and Daniel H. Hill, 1989, "Assessing the Quality of Household Panel Data: The Case of the Panel Study of Income Dynamics", *Journal of Business and Economic Statistics*, Vol. 7, No. 3, 441–452.

Haight, Frank A., 1967, *Handbook of the Poisson Distribution*, (New York: John Wiley & Sons, Inc.).

Hansen, Morris H., W. Hurwitz, and M. Bershad, 1961, "Measurement Errors in Censuses and Surveys", *Bulletin of the International Statistical Institute*, Vol 38, 359–374.

Hausman, J., B.H. Hall, and Z. Griliches, 1984, "Econometric Models for Count Data with an Application to the Patents-R&D Relationship", *Econometrica*, 52:4, pages 909–938.

Johnson, Norman L. and Samuel Kotz, 1969, *Discrete Distributions*, (New York: John Wiley & Sons, Inc.).

Katti, S. K. and John Gurland, 1961, "The Poisson-Pascal Distribution", *Biometrics*, Vol 17, 527–528.

McCullagh, P. and J. A. Nelder (FRS), 1989, *Generalized Linear Models*, (new York: Chapman and Hall).

Maddala, G.S., 1983, *Limited-Dependent and Qualitative Variables in Econometrics*, (Cambridge University Press). pages 51–55.

O'Muircheartaigh, Colm A., 1986, "Correlates of Reinterview Response Inconsistency in the Current Population Survey", *Proceedings of the Second Annual Research Conference*, (U.S. Department of Commerce: Bureau of the Census) pages 208–235.

O'Muircheartaigh C. A. and R.D. Wiggins, 1981, "The Impact of Interviewer Variability in an Epidemiological Survey", *Psychological Medicine*, 11, pages 817–824.

Press, W.H., B.P. Flannery, S.A. Teukolsky and W.T. Vetterling, 1986, *Numerical Recipes*, (Cambridge University Press).

Smith, R., 1987. "The SIPP Results of the Reinterview Program for 1985", (Washington, D.C.: Bureau of the Census, June 18, 1987).

St. Clair, J., 1985. "SIPP-1984 Panel: Results of the Reinterview Program for Waves 2 through 4", (Washington, D.C.: Bureau of the Census, July 2, 1985).

Appendix
Maximum-Likelihood Poisson-Pascal Estimation

As noted in the main text, our maximum-likelihood estimator of the Poisson-Pascal is based on the recursions:

$$P(x=0) \equiv P_0 = \exp[-\lambda(1-(1+\delta)^{-\theta})] \qquad \text{A.1)}$$

and

$$P(x=k) = P_k = [\exp(-\lambda)(\delta/(1+\delta))^k/k!]\,[\sum_{j=1}^{\infty}(\theta j)^{[k]}\{(\lambda(1+\delta)^{-k})^j/j!\}] \qquad \text{A.2)}$$

provided by Johnson and Kotz (1969). In our application the summation is continued until the increment falls below 1.0e-6. When λ is large and θ small, as in the case of the response-error data, this convergence can be quite slow. In this case precision errors can accumulate in the expression $\{(\lambda(1+\delta)^{-k})^j/j!\}$. To eliminate this problem we used the 'extended precision' real number representation allowed by TURBO PASCAL 4.0. This number type occupies ten bytes and provides a range from 1.9 x 10E-4951 to 1.1 x 10E+4932 with 19 significant digits. This degree of precision also allows the computation of very precise numeric derivatives and thereby greatly reduces programming complexity.

One might think that the recursion formula:

$$P_{k+1} = \frac{\lambda}{k+1}\sum_{i=0}^{k}[(k+1-i)\pi_k - iP_i] \qquad \text{A.3)}$$

where A.4)

$$\pi_k = \frac{(\theta+k-1)!}{(\theta-1)!\,k!}\,\delta^{\theta}\,(1+\delta)^{-\theta-k}$$

provided by Katti and Gurland would be a better choice for numerical implementation of the Poisson-Pascal. Unfortunately, we were not able to implement this recursion successfully—the predicted probabilities for valid ranges of the parameters were sometimes negative and the did not sum to unity.

This caused us to be somewhat skeptical of both recursions. To test our maximum-likelihood algorithm which employs the Johnson and Kotz recursions, we estimated the model using the data presented in Katti and Gurland for both the *Lepesdeza Capitata* and the *Leptinotarsa Decemlineata* distributions. Table A.1 presents the our results and Katti and

Gurland's Method of Moments results for the former. Our maximum likelihood method provides estimates which better describe the distribution of both life-forms.

Table A.1
Comparison of maximum-likelihood Poisson-Pascal
with Katti-Gurland Method of Moments

Plants	Observed Frequency	Predicted Method of Moments	Predicted maximum-likelihood
0	7178	7185.0	7181.3
1	286	276.0	282.0
2	93	94.5	94.5
3	40	41.5	40.8
4	24	20.2	19.6
5	7	10.4	10.0
6	5	5.6	5.3
7	1	3.1	2.9
8	2	1.7	1.6
9	1	1.0	.9
10	2	.6	.5
11+	1	.3	.7
χ^2 (d.f.)		9.58 (8)	7.88 (8)

A.2 Recursions for Limiting Forms

While by placing restrictions on its parameters we could use the Poisson-Pascal recursions to estimate the parameters of all its limiting forms, this would not be computationally efficient. Instead, therefore, we use special formulae for each of the limiting forms investigated. These recursions are:

Neyman Type A

As noted in the main text the Neyman Type A estimates are based on:

$$P(n_i = 0) = \exp[-\theta'(1 - \exp(\lambda'))] \qquad A.5)$$

and:

$$P(n_i = k]\ (\theta'\lambda'/k)\exp(-\lambda')\sum_{j=0}^{k-1}(\lambda'^{j}/j!)\ P(k\text{-}j\text{-}1) \qquad A.6)$$

Negative-Binomial or Pascal

The results for the Negative-Binomial, or Pascal distributions are based on:

$$P_0 = \delta^{\theta} \qquad A.7)$$

$$P_k = \frac{k + \theta - 1}{k}(1 - \delta)\ P_{k-1}. \qquad A.8)$$

Poisson

The Poisson results are based on:

$$P_0 = \exp(-\lambda) \qquad A.9)$$

$$P_k = (\lambda/k)P_{k-1} \qquad A.10)$$

Table 1
Whether Received State Unemployment Compensation
As Recorded in the Reinterview by How Recorded in Original Interview

	Reinterview			
Original Interview	Blank	1 'Yes'	2 'No'	Total
Blank	1,250	7	59	1,316
1 (Yes)	7	29	3	39
2 No	29	3	172	204
Total	1,286	39	234	1,559

Table 2
Error Rates for the Substantive Reinterview Questions

Question	error Rates (percent)			
	Overall	Procedural	Response	Dual Response
1. Have job?	1.89	0.26	1.63	99.7
2a. Look for job?	2.20	1.28	2.55	36.18
3a. Want job?	2.81	1.54	4.04	31.37
4. Each week?	3.84	1.92	3.11	61.58
9a. U.I. Comp?	3.15	2.76	2.99	13.28
23a. Medicare?	5.13	5.07	*	4.75
24. Food Stamps?	1.93	1.67	.28	91.02
26a. Mcaid now?	3.08	2.44	0.71	90.89
26b. Mcaid B4?	1.68	1.48	*	3.40
27a. Health Ins?	6.78	0.77	6.03	99.23
27e. Via emplyr?	6.00	4.68	2.35	55.93
27f. Emplyr pay?	7.35	4.11	7.62	42.59

*Rate suppressed due to the small number of cases in the denominator.

Table 3
Maximum-Likelihood Estimates of
Poisson-Pascal and Limiting Form Distributions
for Counts of Response and Total Errors

Distribution Parameters	Response Errors	Total Errors
Moments		
Mean	.171	.572
Variance	.171	1.16
Poisson-Pascal		
θ	1.0	71.78
λ	113.6	0.55
δ	1.5e-3	1.5e-2
ln(L)	−752.02	−1567.89
χ-square	0.054	5.25
(d.f.)	(2)	(8)
Neyman A		
θ'	2.5e-3	1.06
λ'	67.76	0.54
ln(L)	−752.02	−1568.07
χ-square	0.034	5.27
(d.f.)	(2)	(8)
Pascal (Negative-Binomial)		
θ''	24.0	.58
δ''	140.8	.73
ln(L)	−752.02	−1579.54
χ-square		21.8
(d.f.)	0.042	(8)
Poisson		
λ'''	.17	.58
ln(L)	−752.02	−1802.43
χ-square	0.051	390+
(d.f.)	(2)	(8)

*Based on Johnson and Kotz (1969) Section 8.2, equations 42.1 and 42.2.

Table 4
Maximum-Likelihood Poisson Regression Estimates of Response errors
(Asymptotic SRS Standard Errors in Parentheses)

	Bivariate		Multivariate	
	Parameter	Log-likelihood	without Income	with income
Constant (Λ_0)	−3.609** (.037)	−775.4	−1.455* (.623)	−1.724** (.603)
Interview Month	−.275* (.132)	−773.2	−.251+ (.132)	−.235 (.130)
Regional Office error Rate	.935** (.322)	−770.8	.980** (.313)	.962** (.313)
Proxy Respondent	.175 (.132)	−774.5	.107 (.146)	.113 (.146)
Odd Relationship to Reference Person	.383* (.161)	−772.7	.109 (.199)	.030 (.203)
Age (decades)	−.485** (.176)	−771.4	−.369+ (.207)	−.215 (.197)
Age-squared (decades-squared)	.047** (.175)		.035+ (.202)	.205 (.197)
Education	−.044* (.019)	−772.7	−.042* (.021)	−.020 (.022)
Whthr Female	.098 (.123)	−775.0	.080 (.128)	−.071 (.137)
Whthr Black	.162 (.203)	−775.0	.140 (.209)	.128 (.203)
Income ($100's)	−.827** (.211)	−766.5		−.701** (.241)
ln(likelihood) (d.f.)			−761.5 (10)	−757.8 (11)

+significant at the 10% level.
*significant at the 5% level.
**significant at the 1% level.

Table 5
Maximum-Likelihood Neyman Type A Regression Estimates
for Total Errors
(Asymptotic Standard Errors in Parentheses)

	Bivariate		Multivariate	
	Parameter	Log-likelihood	without Income	with income
Constant (Λ_0)	1.063** (.069)	−1568	1.050* (.452)	.863* (.452)
θ'	.542** (.058)		1.008** (.070)	.997** (.071)
Interview Month	−.200* (.094)	−1566	−.191* (.095)	−.181+ (.095)
Regional Office error Rate	.969** (.238)	−1559	1.017** (.236)	1.010** (.235)
Proxy Respondent	.179+ (.094)	−1566	.128 (.110)	.140 (.108)
Odd Relationship to Reference Person	.401** (.111):	−1562	.254+ (.145)	−.187 (.146)
Age (decades)	−.505** (.128)	−1557	−.333* (.152)	−.215 (.150)
Age-squared (decades-squared)	.056** (.012)		.040** (.015)	.029+ (.015)
Education	−.048** (.014)	−1562	−.026+ (.016)	−.009 (.017)
Whthr Female	.005 (.090)	−1568	.−032 (.095)	−.158 (.102)
Whthr Black	.279* (.140)	−1566	.266+ (.147)	.258+ (.147)
Income ($100's)	−.695** (.142)	−1556		−.585** (.175)
ln(likelihood) (d.f.)			−1537 (11)	−1532 (12)

+significant at the 10% level.
*significant at the 5% level.
**significant at the 1% level.

FIGURE 1
SIPP Reinterview Questionnaire Flow

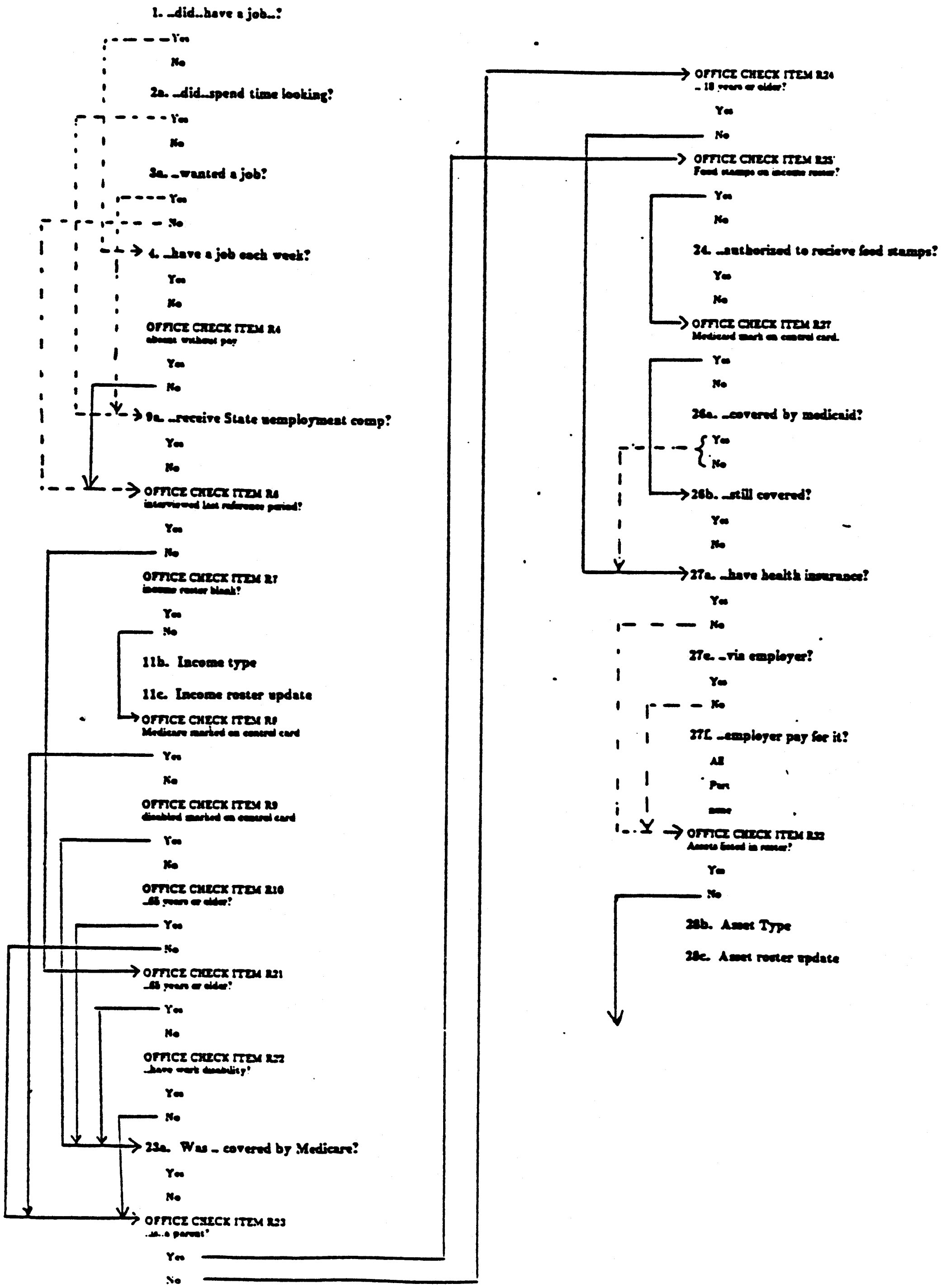

Figure 2
Actual and Theoretical Distributions of Response Errors

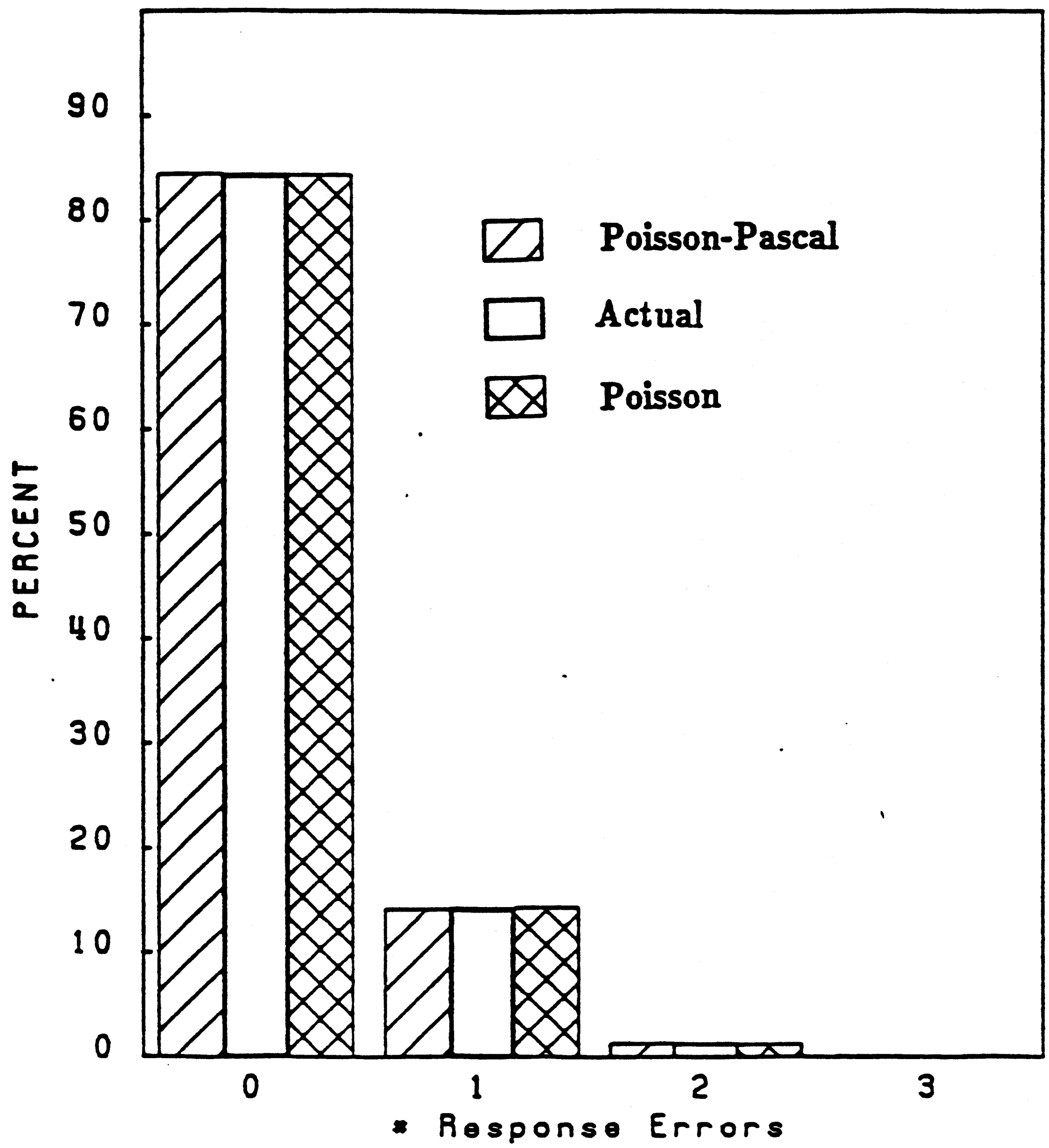

Figure 3
Actual and Theoretical Distributions
of Total Errors

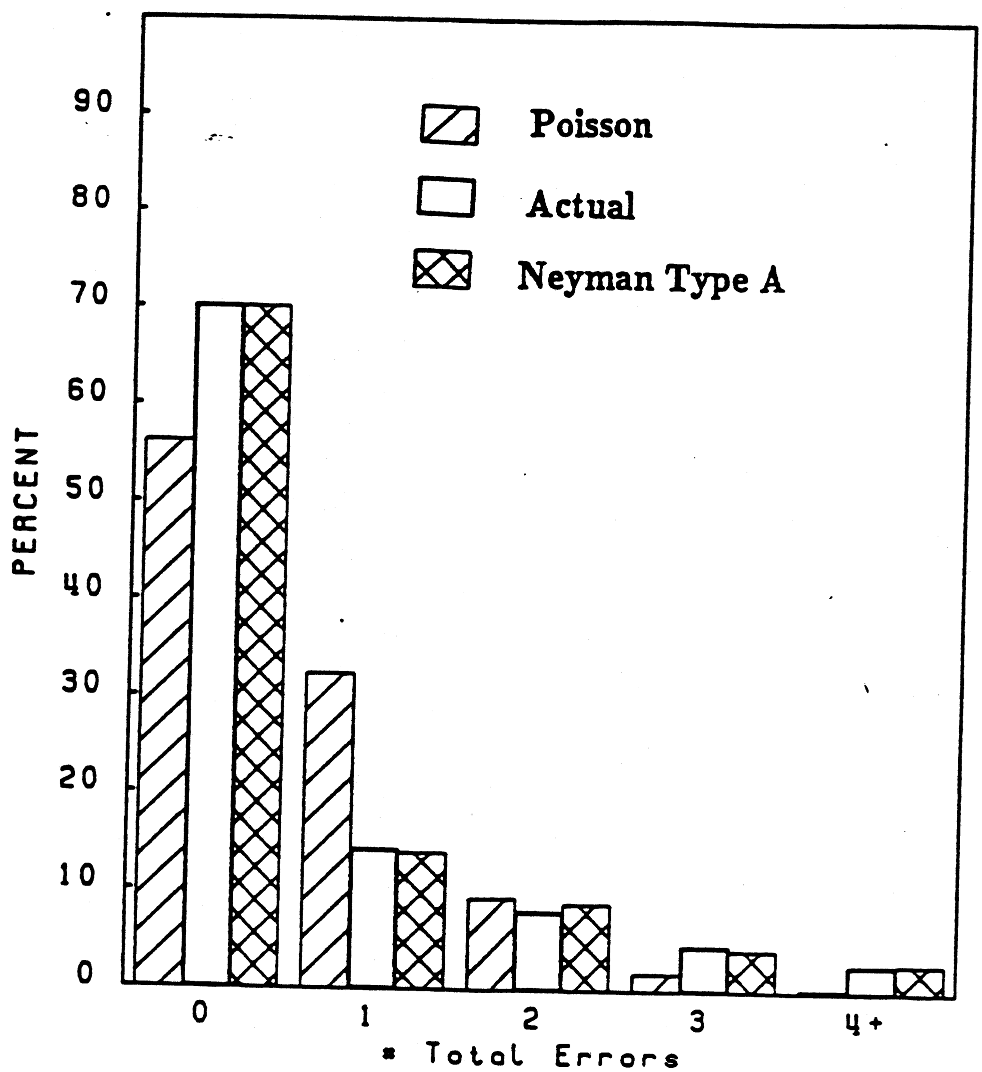

Figure 4
Neyman Type A and Component Poissons
Total Errors