

## Miscellanea

Under the heading Miscellanea, essays will be published dealing with topics considered to be of general interest to the readers. All contributions will be refereed for their compatibility with this criterion.

# Self/Proxy Response Status and Survey Response Quality

## A Review of the Literature

*Jeffrey C. Moore<sup>1</sup>*

**Abstract:** Three decades of research have not produced conclusive evidence of consistent response bias or response error variance differences due to self/proxy response status. The net nonresponse effect may also be close to zero due to compensating effects for the various components of nonresponse. The main cause of the lack of evidence is the methodological shortcomings of much of the research which purports to address the self/proxy issue. In addition, the few methodologically sound studies – most importantly, those which control potential self-selection biases, and whose subject matter makes the

self/proxy distinction appropriate – in general have produced no effects or conflicting effects (or, in the case of nonresponse, compensating effects). However, lack of convincing evidence of quality differences is not synonymous with convincing evidence of no quality differences. Until more data are gathered, the conclusion that self and proxy survey responses are of equivalent quality must remain tentative.

**Key words:** Self/proxy response; response quality; response bias; response error variance; nonresponse; survey design.

### 1. Introduction

Survey research involves many compromises, of which sampling is perhaps the most fundamental. Sampling forces the survey designer to accept reduced estimate precision in exchange for cost and effort feasibility. Other compromises are more subtle. For example, clustered sample designs sacrifice some of the information value of individual responses for enhanced data collection effi-

<sup>1</sup> Center for Survey Methods Research, U.S. Bureau of the Census, Washington, D.C. 20233, U.S.A. The views expressed herein are the author's and do not necessarily represent official views or policy of the Census Bureau. I thank Kent Marquis, Dan Hill, and Betsy Martin for their many helpful and encouraging comments on early drafts of this paper. The final paper has also benefited from the comments of three anonymous *JOS* reviewers.

ciency. Questionnaire design decisions must balance the survey designer's desire for complete information against a reasonable respondent burden. In retrospective surveys, decisions about the length of the reference period attempt to balance the advantages of a short reference period (presumably, reduced memory errors) against a longer period in which the naturally greater number of target events serves to improve estimate precision. Follow-up efforts to reduce nonresponse push against budget and schedule constraints.

In addition to the design decisions they share with all sample surveys, surveys which collect data on each eligible member of each sampled household require one more key design decision: how much effort to expend gathering data *on* individuals *from* each individual himself or herself. The solution to this problem has always varied from survey to survey, even within the same survey organization. For example, the Census Bureau's four major continuing household sample surveys – the Current Population Survey (CPS), the National Health Interview Survey (NHIS), the National Crime Survey (NCS), and the Survey of Income and Program Participation (SIPP) – use four very different sets of procedures to collect data on each household member. As the costs of sample surveys have increased, however, there seems to be a trend to rules more accepting of proxy responding. And yet, despite at least three decades of concern about the effects of self/proxy response status on data quality, survey designers have little evidence to justify the use of more costly procedures or to defend the quality of data obtained less expensively.

The purpose of this paper is to review the literature for evidence on the data quality effects of self versus proxy response status. The central question can be stated in general terms as follows:

For survey items about individual A, are there systematic differences in quality between the responses obtained directly from A versus those obtained from some other respondent who is reasonably likely to be informed?

The common wisdom about self/proxy effects is that the best information about person A will come from person A directly (Sudman and Bradburn (1974); Roshwalb (1982); Mathiowetz and Groves (1983)). There are, of course, recognized exceptions. Proxy responding is generally permitted – if not required – for children and for those too mentally or physically infirm to respond. But this review is concerned with the “standard” proxy situation, in which an eligible and capable adult for some reason does not self-respond. The generally accepted notion is that response quality suffers to the extent that such persons do not respond for themselves. In its common form, then, the core question of this review is explicitly directional:

Do survey data suffer in quality when eligible sampled persons do not respond for themselves? To what extent is quality sacrificed when a survey designer opts for rules which permit proxy response?

## **2. Methodological Considerations**

The assumption of quality differences favoring self-response has intuitive appeal, although in certain instances the opposite case can be made, such as when the survey subject matter may evoke self-presentation pressures (e.g., Berk, Horgan, and Meyers (1982)). However, the standard assumption also seems to draw support from the self/proxy literature. Unfortunately, a substantial portion of this literature only appears to address the data quality implications of alter-

native respondent rules; close inspection often reveals conceptual and methodological shortcomings which render conclusive judgments about self/proxy quality effects impossible. Much of the research which purports to address the question of self/proxy response quality differences falls short on at least one of three important criteria: (1) survey subject matter, (2) self/proxy status control, and (3) quality assessment.

### *2.1. Survey Subject Matter*

The most obvious criterion for the examination of self/proxy effects is that the survey inquiry must refer unambiguously to an individual. If this criterion is not satisfied, the self/proxy distinction is not meaningful. There is an extensive "pseudo proxy" literature which probably contributes to the consensus judgment regarding self/proxy effects. For example, husbands and wives have been found to provide very different accounts of their relative influence in household purchase decisions (Ferber (1955a)), and of family economic characteristics in general (Ferber (1955b)). Spouses' reports of frequency of intercourse (Levinger (1966)) and of other "shared experiences" (Mudd, Stein, and Mitchell (1961)) show large discrepancies. Parents and children disagree about past childrearing practices (Radke (1946); Kohn and Carroll (1960)).

There are two main problems with this literature. First, although respondent discrepancies seem to be the general rule, the data are by no means uniform; many studies have found substantial agreement between respondents. (See for example, Neter and Waksberg (1965) on household expenditures; Haberman and Elinson (1967) on family income; Kinsey, Pomeroy, and Martin (1948) and Rutter and Brown (1966) on sexual behavior and other relationship variables.)

The second and more critical problem is that the use of the terms "self" and "proxy" is simply not appropriate for these topics. Many survey questions are about the past behaviors, life events, and current circumstances and characteristics of people not as individuals but as participants in some collectivity – spouses, families, households. While reporting consistencies and discrepancies among collectivity members pose interesting questions for survey methodologists, it is not legitimate in these investigations to identify any particular member as a "self" respondent. The literature searches carried out for this review captured studies of this type with some frequency, suggesting that conventional wisdom about self/proxy differences may be contaminated by irrelevant research findings.

### *2.2. Self proxy treatment control*

The second criterion concerns proper research design. Unfortunately, a common "design" in self/proxy research is no design at all – a survey is conducted, some people respond for themselves and others are responded for by proxy, the responses of the two naturally-occurring groups are compared, and conclusions are drawn about the effects of response status on response quality. Without strong assumptions, however, such conclusions are not justified.

Definitive research evidence can only derive from studies which can dispense with reasonable competing explanations for observed effects. Studies of naturally-occurring self/proxy effects are open to the possibility that observed differences are a result of self-selection biases and do not necessarily indicate response quality problems for one or the other group. Thus, the typical finding in health surveys of more frequent reporting of health conditions, doctor visits, hospital stays,

etc. for self-responders than for those responded for by proxy (e.g., Horvitz (1952); Enterline and Capt (1959); Linder (1959) (cited in Cartwright (1963)); Haase and Wilson (1972); Berk et al. (1982)) may simply reflect the greater likelihood of finding the less healthy household members at home when an interviewer calls. (Haase and Wilson (1972), Kovar and Wright (1973), Berk et al. (1982), and others who have identified this effect have noted the possibility that true sample differences account for the reporting differences of self and proxy respondents.)

Although procedures to assess response quality (see Section 2.3.) make tantalizing additions to uncontrolled treatment studies, they do not render the evidence on response status effects any more conclusive. Such studies still cannot discount self-selection sample bias as a possible explanation for observed quality differences. Thus, studies showing more complete survey reporting of medical conditions for self-respondents than for proxies as judged against a subsequent medical examination (Commission on Chronic Illness (1957); Elinson and Trussell (1957); Krueger (1957)), for example, or more accurate income reporting according to a match to administrative records (Kilss and Alvey (1976)), are fundamentally uninformative because they fail to address the key question: if the original proxy group had responded for themselves, would their reports have shown any greater correspondence to the validating data?

Of course, this difficulty afflicts all uncontrolled self/proxy studies, regardless of the direction of their results. Thus, the conclusion of Berk et al. (1982) that proxies produce *better* quality reports of stigmatizing physical conditions also goes beyond the available data. Unless self/proxy status is controlled, the conclusion that quality suffers – or is enhanced – as a result of proxy status rests on an untested assumption of self and proxy sample comparability.

### 2.3. Response quality assessment

The third methodological criterion is a concrete assessment of data quality. A long-standing tradition in survey nonsampling error research is that a definitive evaluation of response quality cannot occur without reference to a “true value” (Hansen, Hurwitz, Marks, and Mauldin (1961)), or “some external criterion” (Sudman and Bradburn (1974)). This approach to data quality, then, is concerned with the deviation of individual survey responses from some external standard of truth.<sup>2</sup> Aggregated across a set of responses, these deviations can be used to assess the total error associated with self and proxy response, and the extent to which response errors under the two conditions are systematic or random.

Response quality is best evaluated through a comparison of individual survey responses with some independent, external criterion, such as existing records, or an objective, non-survey-based measurement of the same phenomenon. Since a well-designed validity assessment is difficult to execute, many self/proxy studies have taken the easier course of simply comparing the aggregate responses of the two respondent groups. Such studies typically adopt a “more-is-better” assumption (or the reverse for socially desirable subject matter), and occasionally even a “self-is-true-therefore-proxy-differences-indicate-error” assumption. The latter simply assumes what should be a matter for objective inquiry. The former probably derives from the results of one-directional or partial record check designs, in which only a limited range of survey reports is validated. As Marquis (1983) and Marquis, Duan, Marquis,

<sup>2</sup> An important corollary of this definition of quality is that the subject matter of eligible studies must offer at least the potential of external verification, which excludes survey measures of “subjective phenomena” (Turner and Martin (1984)).

and Polich (1981) point out, such designs are guaranteed to produce apparent bias estimates in only one direction.

Studies lacking a direct assessment of response quality can still yield useful information about self/proxy effects. If controlled response status treatments produce reliable self/proxy report differences, then we may infer that a stable response quality effect exists, even though the critical details of that effect (What are the directions and magnitudes of the biases? Which group's data are better?) must await additional research. However, a finding of no consistent self/proxy report differences does not necessarily indicate quality equivalence. The systematic "noise" associated with each response status may be comparable, but the random noise may differ greatly.

Because quality assessment is difficult to implement (and sometimes seemingly impossible), researchers with otherwise appropriately designed studies have often fallen back on more indirect quality indicators. The most frequent of these are the various dimensions of response completeness – item nonresponse, person nonresponse, and household nonresponse. Studies which consider the effects of self/proxy status on these (and other) indirect quality indicators are also included in the review.

### 3. Literature Review

Survey procedures which attempt to maximize self-response are more expensive than those which are more tolerant of proxy responding, and the difference can be enormous for large survey programs. By not requiring self-response, total CPS costs are effectively reduced by about 12 percent (Deighton (1967)) – or \$150 000 per month at current rates. Estimates for other surveys

have indicated that maximum self-response rules would increase costs by 5 percent for a comprehensive income survey (Kulka (1982)), 17 percent for a health survey (Kovar and Wright (1973)), and up to 30 percent for a Canadian labor force survey (Singh and Tessier (1975)).

These studies leave little doubt that self-response rules are the more expensive option. Do they buy better quality data? The remainder of this paper reviews the evidence on self/proxy response quality differences for three dimensions of response quality – response bias, response error variance, and nonresponse. In keeping with the discussion in Section 2, I consider here only those studies which meet the appropriate subject matter and self/proxy control criteria. Although quality assessment procedures are desirable, otherwise well-designed studies lacking this component are still informative, and so are also included in this review.

#### 3.1. Response bias effects

##### 3.1.1. Report level differences

A large portion of the self/proxy literature consists of studies which control the assignment of respondents to self or proxy response but which lack a direct assessment of response quality. These studies offer little evidence of consistent differences in overall report levels attributable to response status, and thus suggest no major differences in the extent of systematic error.

Crime surveys are a possible exception to this general rule. Turner (1972) presents results of the only known controlled self/proxy study in this area. Each household in a large sample ( $n=10\ 000$ ) was randomly assigned to either a household respondent or a self-response treatment. For all eight crimes examined – strong-arm robbery, armed robbery, robbery attempts, aggravated assault,

simple assault, assault attempts, rape, and rape attempts – the self-response treatment produced more incident reports. Although Turner presents no significance tests, the differences are substantial. Again, these results only indicate that the biases associated with self and proxy reporting of criminal victimization are different, and not that one status elicits better data than the other.<sup>3</sup>

Similar investigations have been carried out in labor force surveys, although here the evidence suggests no response bias differences. An experiment comparing self-response to the standard CPS household respondent procedure produced very small (and probably nonsignificant) differences in labor force participation and unemployment rates (Deighton (1967)). Williams (1969) and Jones and Aquilino (1970) report results of another experiment in the CPS, involving dual interviews with two designated respondents, each of whom reported for self and all other household members. The experiment yielded near perfect agreement between self-reporters and proxies on labor force participation and unemployment rates.

A more complex CPS experiment was conducted a decade later, including a comparison of the standard household respondent procedure with a designated household respondent and a maximum self-response procedure. There were scattered significant interactions involving the respondent treatment variable. However, keeping the other experimental factors constant, there were no significant response status differences in esti-

mated unemployment rates (Cowan, Roman, Wolter, and Woltman (1979); Roman and Woltman (1980)).

Martin and Butcher (1982) present results of a self/proxy experiment involving U. K.'s Labour Force and National Dwelling and Housing Surveys, in which independent interviews were conducted with both proxies (usually wives) and self-respondents about the latter's labor force activity and related issues. Martin and Butcher report very high levels of agreement between self and proxy reports across a wide range of topics, including labor force status, type of occupation, transportation to work, hours worked last week, age, completed education, etc.

Self/proxy research has a long history in the health survey area. In perhaps the earliest experimental examination of self/proxy treatment differences, Cartwright (1957), in a small pilot study, compared husbands' self-response health reports against those of their wives acting as household respondents. The average number of reported illnesses per husband was three times greater under the self-response treatment than under the proxy treatment (2.5 versus 0.8), and the proportion reporting no illnesses was about one-sixth as great (8% versus 49%).

Most early health studies, however, report no self/proxy report differences or inconsistent differences. Enterline and Capt (1959) randomly assigned adult males not at home on an initial visit to an immediate proxy interview or a self-response followup. They found no significant treatment differences for any of the ten specific chronic conditions examined, and the average number of conditions reported per person was identical for the two procedures. Items assessing parents' histories of heart disease also showed no consistent differences. Proxy estimates of daily fat intake were marginally higher than those obtained via self-response, but items concerning age, weight, stature, weight gain, and

<sup>3</sup> The quality implications of differing report levels are particularly ambiguous in the criminal victimization area. On the one hand, the common assumption is that victimization is generally under-reported (thus, more reporting means better reporting). On the other hand, there is also great concern about "forward telescoping," in which incidents which occurred prior to the stated reference period are reported as having occurred within the reference period (thus, more reporting may mean worse reporting).

recent medical care showed no significant differences.

Similarly, in an early experimental pretest of the NHIS, adults in a sample of households were randomly assigned to either a strict self-response treatment, or a treatment which accepted proxy responding (Nisselson and Woolsey (1959)). The overall illness rate was significantly greater for the self-response treatment than for the proxy treatment, although the results are quite inconsistent across sex, age, and illness categories. Reported days of disability were consistently lower in the self-response group, although with only scattered significant effects. The authors' main conclusion, however, is that "the sampling and response variability in the data are too high to permit definitive conclusions as to possible biases in the use of a household respondent" (Nisselson and Woolsey (1959, p. 72)).

Kovar and Wright (1973) carried out a similar experiment, on a much larger scale, comparing a maximum self-response treatment with a more lenient "accept-proxy" treatment. For most items there appear to be no significant treatment differences,<sup>4</sup> although reporting is fairly consistently higher under self-response conditions. Only two items – "limitation of activity" reports, and reports of recent doctor visits – show a significant treatment effect, with the self-response treatment producing significantly higher report levels. Although "limitation of activity" is clearly not identical to Nisselson and Woolsey's (1959 op. cit.) "days of disability," there does appear to be some inconsistency in these two sets of results.

As part of an investigation of increased use of the telephone for sample selection and interviewing in the NHIS, a recent experiment compared a randomly designated household respondent procedure with a self-selected "knowledgeable" household respondent procedure. Mathiowetz and Groves (1983) present treatment differences for the same set of indicators as Kovar and Wright (1973 op. cit.). Their findings, however, are quite the opposite of the earlier research. Most of the differences, including the only statistically significant effect, are in the direction of higher reported levels for the knowledgeable respondent (proxy) condition than for the self-responses from the randomly designated respondents. Within the random respondent treatment there was also a consistent trend for higher reporting levels among proxies. There is some evidence of the more "usual" trend for measures involving a longer recall period, but Mathiowetz and Groves summarize the results as showing an "overall tendency... directly counter to previous beliefs about self vs. proxy reports" (p. 96).

Several studies in the health area have controlled selection bias through a reinterview design, in which respondents originally interviewed by proxy are subsequently interviewed in person. The earliest of these occurred in conjunction with the California Health Survey. First, in a small pilot survey, 118 original proxy respondents were administered a self-response reinterview approximately two weeks after the initial interview. The reinterview was conducted without reference to the original interview, using the identical questionnaire. In general, the original proxy interview produced fewer reports of medical conditions than did the self-response reinterview. Discrepancies were substantial in both directions, however; each survey elicited many condition reports which had not been included in the other (California Department of Public Health (1957); Mooney (1962)).

<sup>4</sup> The authors contend that three of the two-week recall items show significantly higher reporting levels for the self-response treatment. However, their use of one-tailed significance tests is questionable, since two of the nonsignificant differences are in the direction of *more* reporting for the accept-proxy treatment.

A substantially larger reinterview program was also carried out in the subsequent main survey. For all of the health indicators examined – chronic conditions, acute conditions, restricted activity days, bed disability days – the self-response reinterview yielded a substantial increase in medical events and conditions over the original proxy reports. However, this increase is not totally attributable to self/proxy response status, since the reinterview also produced increased reporting among a sample of original self-respondents. The percentage increase from original interview to reinterview was greater for the original proxy group than for the original self-respondents, suggesting that response status contributed something to the difference (California Department of Public Health (1957); Mooney (1962)).

Koons (1973) has analyzed similar data from several years' reinterviews in the NHIS. His results also show increased reporting of health events and conditions in a self-response reinterview as compared to an original proxy interview, and the increases are typically larger than the interview-to-reinterview increases for those responding for self in the original interview. Thus, the results of these reinterview studies consistently suggest that the biases associated with self and proxy reporting of health conditions may differ, although differences between interview administrations may far outweigh any self/proxy differences.

In another health survey involving a type of reinterview design, Kolomel, Hirohula, and Nomura (1977) independently interviewed both members of 300 couples (mostly spouses) about each person's smoking, drinking, and diet behavior. The authors report great consistency between self and proxy reports, leading them to conclude that, for such inquiries, proxy reports are equal in quality to self-reports.

Only within the last few years have researchers begun to examine the effects of respondent status on income reporting. The 1979 panel of the Income Survey Development Program (ISDP) was the final major pretest in preparation for the Survey of Income and Program Participation (SIPP). This panel included an experiment on respondent rules, in which each sample household was randomly assigned to a maximum self-response treatment or an "accept-proxy" procedure.

Ferber and Frankel (1981) examine household ownership rates for 17 asset types (e.g., bonds, savings accounts, stocks, rental property, royalties, etc.) by respondent rule procedure. Differences in ownership rates are generally small, but consistently higher under the maximum self-response treatment. However, in a general summary of this test, Kulka (1982) concludes that the self/proxy response treatments produced no significant differences in asset reporting.

Income reciprocity rates for both earned income and unearned government transfers also appear to have been unaffected by respondent rules. Kaluzny (1981a, 1981b) reports results from the same ISDP panel which show no significant differences for 13 income types. For respondents who reported any income, reported income amounts also did not differ.

Another self/proxy comparison in the ISDP is possible due to a special procedure in one wave of the 1979 panel – a self-response followup survey of students not living at home (and thus originally interviewed by proxy). Roman (1983) presents a comparison of self and proxy reports for 167 students who were successfully followed. The results suggest a higher rate of receipt of wage or salary income from the self-reports (66% versus 52% for proxy interviews), and interest income receipt (78% versus 66%), but no



significant differences in amount of educational assistance reported.

Martin and Butcher's (1982) self/proxy experiment, described earlier, also finds considerable consistency between self and proxy income reports. Of course, the proportion of reports "in agreement" depends on the definition of agreement: 81% of self-respondents and proxies placed gross weekly income in the same £50 category, 46% in the same £10 category, and less than a third in the same £1 category. Reported mean income differed only trivially, however – £74.8 per week for self-respondents versus £72.4 for proxies.

Hill (1987) offers a promising new technique for research on self/proxy issues. Using a modeling approach to control self-selection bias in uncontrolled self/proxy research, his results suggest that proxy reports of earnings income are substantially positively biased relative to self reports. Hill's work clearly underscores the dangers of drawing simple inferences from uncontrolled self/proxy studies. In this case the research bias is a misleading equivalence of mean self and proxy income reports (prior to the application of statistical controls) when – due to such sample differences as hours worked per week – there should be a difference.

Evidence on self/proxy report differences can also be found in Kinsey et al.'s (1948) investigation of sexual behavior, in which they conducted separate interviews with each spouse in a group of husband/wife pairs. The authors compare spouses' reports on 32 items, most of which involve relationship issues for which the self/proxy distinction is not appropriate. For seven of the items, however, one member of the pair is clearly the object of inquiry: husband's education, age at marriage, and occupation; and wife's education, age at marriage, number of abortions, and percentage of coitus with orgasm. None of these items shows a significant difference between the mean responses of hus-

bands and wives.

Thus far, an examination of self/proxy report level differences from controlled treatment studies yields evidence of consistent effects only for reports of criminal victimization – and the latter is based on only one study. Labor force participation and income surveys have almost universally shown no significant treatment effects. (Hill's (1987) technique, which does indicate self/proxy bias differences for income reports, may now open up past and future uncontrolled treatment studies for more informative analysis.) This is also the most common result in health survey experiments, although some significant effects – albeit contradictory from one study to the next – have been found in the health area. The most appropriate general conclusion to draw from this type of research is that the weight of the evidence does not indicate consistent differences in the reporting levels of self and proxy respondents.

### 3.1.2. Response quality differences

Self/proxy differences in report levels at best only suggest differences reporting quality. In order to inform the issue with confidence, controlled studies need to include a direct assessment of response quality. Appropriately designed record check studies provide the most incontrovertible evidence of self/proxy response quality differences, but such studies are rare, and thus research employing more indirect indicators of response bias will also be examined.

Turner's (1972) crime survey respondent rule research described previously did not include an independent validity check. However, compared to self-respondents, household (proxy) respondents showed a greater tendency to distribute their victimization reports unevenly over the twelve-month reference period. Both self and proxy respondents tended to report more incidents

in the most recent six months of the reference period; for most incidents, however, this effect was more pronounced for the household respondent procedure. For all incidents combined, this treatment produced 50% more reports in the first half of the reference period than in the second half; for the self-response treatment the comparable figure was 41%. These differences suggest that for at least one component of response quality – the accurate dating of crime events, or their more complete recall – proxy responses may be more biased.

Deighton's (1967) comparison of a maximum self-response and a standard household respondent procedure indicated no reduction in the typical CPS "month-in-sample" bias (see Bailer (1975)) with a self-response procedure. In fact, the rotation group differences in the "in labor force" and the "unemployed" categories are more pronounced in the self-response treatment, although there are no statistically significant effects. Similarly, Aquilino (1971) presents results from an experimental panel of the CPS in which the data for all adults were obtained by self-response in the first month, and by both self and proxy response in the second month. The results show equivalent change in labor force classification from the first to the second month for self-self and self-proxy reports. Thus, although their quality assessment procedures are indirect, neither of these two investigations yields evidence of response bias differences between self and proxy labor force reports.

This review uncovered only two studies – both health surveys – in which controlled self/proxy response status was combined with an independent, comprehensive assessment of data quality. First, Cobb, Thompson, Rosenbaum, Warren, and Merchant (1956) summarize results of a three-phase investigation involving: (1) a household health interview, the first question of which asked whether any

household member had arthritis or rheumatism, and if so, who; (2) an individual (self-response) interview with a subsample of persons from the household sample, using a questionnaire devoted solely to arthritis and rheumatism; and (3) a medical examination of a subsample of the self-respondents to detect either current symptoms or a history of arthritis or rheumatism.

Of the 707 persons selected for all three phases of the study, only 429 (61%) provided complete data, with most of the attrition due to refusal of the medical examination. However, the authors assert that "only minor differences were found in examination rate by age, sex, income group, history of arthritis, joint pain, or joint swelling and many other variables." and that the differences "are insufficient to affect the conclusions" (p. 135).

The medical examination placed each person in one of four groups: (1) definite arthritis or rheumatism; (2) symptoms of arthritis or rheumatism but no definite diagnosis; (3) not classifiable in (1) or (2) but without certainty that the person had never had arthritis or rheumatism; and (4) no suspicion of arthritis or rheumatism. (The original paper presents a five-category diagnostic scheme; for simplicity, two of the original categories – "classical arthritis" and "definite arthritis" – are combined here in (1).) Table 1 compares the arthritis or rheumatism prevalence estimates from the household and individual interview reports with the physician's diagnosis. Regardless of how one simplifies the medical diagnosis into a present/absent scheme – that is, whether (1) alone indicates the presence of arthritis or rheumatism, or (1) or (2), or any category but (4) – the bias difference between the reports obtained under the two procedures is very small.

Thompson and Tauber (1957) report results of a similar three-phase investigation of heart disease, involving a household inter-

Table 1. Estimates of arthritis or rheumatism based on a household (proxy) interview, an individual (self) interview, and a medical examination. Percent (Cobb et al. (1956))

Survey Prevalence rate		Examination Prevalence rate		Net bias (Survey – Examination)	
Household	Individual			Household	Individual
34.0	37.5	(1)*	34.7	– 0.7	+ 2.8
34.0	37.5	(1,2)*	51.0	–17.0	– 13.5
34.0	37.5	(1,2,3)*	76.7	–42.7	– 39.2

\* Positive diagnosis criteria (see text).

Table 2. Estimates of heart disease based on a household (proxy) interview, an individual (self) interview, and a medical examination. Percent (Thompson and Tauber (1957))

Survey Prevalence rate		Examination Prevalence rate		Net bias (Survey – Examination)	
Household	Individual			Household	Individual
24.7	33.0		33.0	–8.3	0

Table 3. Estimates of heart disease based on a household interview, an individual interview, and a medical examination for persons interviewed for self in the household interview. Percent (Thompson and Tauber (1957))

Survey Prevalence rate		Examination Prevalence rate		Net bias (Survey – Examination)	
Household	Individual			Household	Individual
24.4	34.2		35.8	–11.4	–1.6

view, an individual interview, and a medical examination. Only about half of those sampled for the individual interview and examination actually completed all three phases of the study. Table 2 summarizes Thompson and Tauber’s results for persons who were responded for by proxy in the household survey and who subsequently responded for themselves in the individual interview. This table appears to confirm the traditional assumption of proxy underreporting; addi-

tional evidence in Table 3, however, suggests another explanation. Table 3 presents the household and individual interview results for persons who responded for themselves in the initial household interview. The household interview bias attributed to proxies in Table 2 is equally apparent in the data summarized in Table 3, where no proxies are involved. The “proxy bias” interpretation of Table 2 is not justified, since the same bias difference occurs

among those who self-responded in both interviews.

Perhaps the household interview suffered because it used only a single, global question to elicit reports of heart disease, whereas the individual survey was “oriented solely toward diseases of the heart” (Thompson and Tauber (1957, p. 1131)). Or perhaps the 18-month delay between the household interview and the medical examination (versus the week or so delay for the individual interview) was at fault; longer delay would increase the likelihood of real change in health status, and real change would almost certainly masquerade as underreporting. Other possibilities may also exist, but whatever the cause of the Thompson and Tauber findings, it is not likely to have been “proxy bias.”

Thus, inspection of the relevant research does not reveal strong or consistent evidence to support the notion that proxy data are in general more biased than self-response data. The results of a single study suggest a slight tendency for self-respondents to distribute crime incident reports more evenly across a 12-month reference period than do proxies. However, two investigations of labor force reporting show no bias differences due to response status, and two health studies, which included direct checks of response validity, also fail to support the assumption that self-reports are less biased than proxy reports.

3.2. Response error variance effects

Another important dimension of data quality is the extent to which respondents reply accurately, regardless of the direction of their errors. A procedure may yield unbiased estimates without producing a single accurate reply (as long as the errors are perfectly compensating), and a more biased procedure may produce a greater number of accurate replies than a less biased one. If first-order point estimates are all that is of interest, then bias is the only component of data quality that need be considered. If, however, the higher-order aspects of the data are of interest (e.g., transition estimates, multivariate relationships, etc.), then response error variance is also important.

As was the case for response bias, sound research on response error variance is rare and often employs only indirect measures of quality. In fact, the clearest data on response error variance effects are from the two health studies described in the previous section.

Table 4 summarizes the Cobb et al. (1956) arthritis and rheumatism results in terms of the gross accuracy rate – the proportion of survey responses which agree with the medical examination – for the household (proxy) and individual (self) interviews. Regardless of how the medical diagnosis of arthritis or rheumatism is defined, the gross accuracy rate difference between the household and the individual interview is trivial.

Table 4. Proportion of household (proxy) and individual (self) interview responses regarding arthritis or rheumatism agreeing with a medical examination. Percent (Cobb et al. (1956))

Medical examination categories indicating a positive diagnosis (see text for explanation)	Gross accuracy rate	
	Household	Individual
(1)	71.8	71.6
(1,2)	73.2	73.2
(1,2,3)	55.9	59.9

Table 5. Proportion of household and individual interview responses regarding heart disease agreeing with a medical examination, for all respondents and for those responded for by a proxy in the household interview. Percent (Thompson and Tauber (1957))

Respondent group	Gross accuracy rate	
	Household	Individual
All respondents	77.3	76.8
Proxy in household interview	73.1	72.5

Thompson and Tauber’s (1957) data, summarized in Table 5, point to the same conclusion. The proportion of accurate replies is virtually identical in the household and individual interviews, even restricting consideration to just the subset of respondents for whom the initial household interview was actually a proxy interview.

The previously described experimental panel of the Income Survey Development Program (ISDP) included a respondent rules experiment, comparing a maximum self-response procedure with an accept-proxy procedure. Evaluations of this experiment have employed various indirect indicators of quality (such as the extent of rounding of income amounts, variances of income amounts, and the respondent’s use of records to assist accurate reporting of income amounts) related to the random error dimension of quality.

Kaluzny (1981a, 1981b) presents results indicating some tendency for more reporting of rounded (i.e., divisible by 5) income amounts under conditions more tolerant of proxy response, but the differences are not consistent across all income types nor within the same type across survey waves. Differences in the variances of reported income amounts are also inconsistent. Income amount variances are generally lower for the self-response treatment in wave 1 of the 1979 ISDP panel, but the wave 2 results show an equal number of differences in both direc-

tions. The only clear evidence in the ISDP of a quality difference favoring the self-response procedure is in the respondent’s use of records (Kaluzny (1981a); Vaughan (1980)).

As noted in the preceding section, a recent paper by Hill (1987) uses a modeling approach to control self-selection bias in uncontrolled self/proxy research. Hill’s results suggest substantially greater response variance for proxy reports of earned income than for self-reports. However, this difference appears due to a few extreme cases; removing these outliers reverses the original difference, resulting in significantly lower variance for proxy reports. Martin and Butcher (1982) report a slightly higher variance for proxy income reports in their self/proxy, dual interview study, although for respondents of higher “social class” or educational attainment this trend is reversed.

Once again, the limited research evidence fails to support the hypothesis that proxy data are inferior to data obtained by self-response – specifically, that they are beset with greater numbers of inaccurate replies. Two health studies show virtually identical levels of response accuracy for self and proxy reports. Attempts to find quality differences (with several different quality indicators) in income reporting show only weak and inconsistent effects, with the exception of the respondent’s use of records to assist accurate recall.

### 3.3. *Nonresponse effects*

Response status may also affect data quality through its effects on response completeness – item nonresponse, person nonresponse, or household nonresponse. Results from two labor force surveys suggest that self-response rules may produce less complete data. Deighton (1967) reports a slightly (but not significantly) higher household noninterview rate with a self-response procedure – 6.2%, versus 5.9% for the standard household respondent treatment – and some person noninterviews in interviewed households (0.6%) where the standard treatment had none. Singh and Tessier (1975) report even more dramatic results in an experimental self-response panel of Canada's Labour Force Survey. The self response panel had a household noninterview rate of about 11%, versus only 6% in the household respondent parent survey.

Kovar and Wright (1973) found no differences in household nonresponse between an experimental self-response treatment and the standard (accept-proxy) NHIS response rules. However, they do report a very small increase in person nonresponse (of about 1%) under self-response conditions. Kovar and Wilson (1976) suggest that this latter effect may only apply to males; regardless, it is probably too small to be of practical significance.

The experiment conducted in the 1979 panel of the ISDP has generated extensive investigations of self/proxy nonresponse effects. For example, Vaughan (undated (a); undated (b)) presents data from the first two (of six) waves of the 1979 panel suggesting slightly higher household refusal and total noninterview rates for the self-response treatment. Olson (1981) corroborates this tendency for all waves of the 1979 panel, but concludes that "attrition differences by treatment are too small to give guidance in the self/proxy decision" for the SIPP itself (p. 1).

The most interesting aspect of these results is that the differences appeared as early as the first survey wave, before respondents had any opportunity to be affected by the self-response procedures. This suggests that noninterview rate differences favoring the accept-proxy procedure might better be attributed to interviewer reluctance to administer the self-response procedure than to any negative reaction from respondents.

Similar slight differences are evident in person noninterview rates within interviewed households. Vaughan (1980) reports a 2.5% noninterview rate with the maximum self-response treatment in the first ISDP wave, versus 1.2% under the standard accept-proxy treatment. Vaughan characterizes this difference as statistically but not practically significant (p.2).

Much attention has been directed toward comparisons of item nonresponse for the two respondent rule treatments in the ISDP. This attention undoubtedly reflects the great concern with nonresponse rates for income items in the traditional major sources of income data in the United States – the decennial census and the March CPS income supplement. Coder (1980) presents first wave nonresponse rates for six income items in the 1979 panel: hourly wage rate, Social Security Income, Federal Supplemental Security Income, pension and retirement income, self-employment income, and rental income. In each case, item nonresponse under proxy conditions is substantially higher than under the self-response rule. Vaughan (1980) reports nonresponse information for six income variables, which also show consistently greater nonresponse under the accept-proxy treatment.

Both Coder and Vaughan used preliminary and unedited data, which may explain why they are at odds with Kaluzny's (1981a, 1981b) later examination of item nonresponse on amounts received from seven

income sources. Kaluzny's results present a much less consistent picture. In wave 1 only four of the seven income types show greater item nonresponse under accept-proxy conditions, and most of the differences are reversed in wave 2. Kulka (1982) summarizes the various investigations as not having demonstrated consistently lower item nonresponse rates under self-response, with the possible exception of hourly wage and Social Security Number reporting. Martin and Butcher's (1982) data, however, support the more usual trend – for all of the labor force and income items examined, proxy item nonresponse (including “don't know” and “uncodeable” replies) exceeded that of the self-respondents.

Although the effects are neither large nor consistent, the research evidence suggests that self/proxy status may have some reliable effects on response completeness. Across subject-matter areas, it appears that self-response procedures produce higher household and person noninterview rates. Overall response completeness may be equivalent, however, since these differences seem to be balanced by lower item nonresponse under procedures which maximize self-response.

#### 4. Summary and Conclusions

This review of the literature finds little support for the notion that self-response survey reports are of generally better quality than proxy reports. In practical terms, this suggests that survey designers should use self-response interviews if they are easily obtainable, but need not undertake extraordinary efforts to maximize self-response. Crime studies may be an exception to this generalization. The existing data should give researchers some confidence that the responses of proxies are comparable in quality to what would have been obtained via self-response with additional effort.

The more obvious conclusion is that there is really not enough evidence to draw solid conclusions. Well designed studies of the self/proxy issue are very rare, and the range of subject matter covered has been limited. Furthermore, only the most basic overall estimates have received any attention. It is quite possible, for example, that self/proxy status may have negligible general effects, and yet may interact with respondent characteristics so as to significantly affect age-specific (or other) estimates. Martin and Biderman (1984) and O'Muircheartaigh (1986) present data which suggest this possibility.

So, this review, too, must conclude with a call for more – and better – research. Too much of the work that has been done has been relatively easy and inexpensive, but has not really advanced our knowledge or provided practical guidance for survey planners. A sound research program would provide the information for more rational decision-making.

#### 5. References

- Aquilino, R. (1971): *Methods Test Phase III: Third Report on the Accuracy of Retrospective Interviewing and Effects of Change in Respondent on Labor Force Data*. U.S. Bureau of the Census memorandum (April 2, 1971), Washington, D.C.
- Bailar, B.A. (1975): The Effects of Rotation Group Bias on Estimates from Panel Surveys. *Journal of the American Statistical Association*, 70, pp. 23–30.
- Berk, M.L., Horgan, C.M., and Meyers, S.M. (1982): The Reporting of Stigmatizing Health Conditions: A Comparison of Proxy and Self-reporting. *American Statistical Association, Proceedings of the Section on Social Statistics*, pp. 506–509.
- California Department of Public Health (1957): *Health in California*. Department of

- Public Health, Berkeley, CA.
- Cartwright, A. (1957): The Effect of Obtaining Information from Different Informants on a Family Morbidity Inquiry. *Applied Statistics*, 6, pp. 18–25.
- Cartwright, A. (1963): Memory Errors in a Morbidity Survey. *Millbank Memorial Fund Quarterly*, 41, pp. 5–24.
- Cobb, S., Thompson, D.J., Rosenbaum, J., Warren, J.E., and Merchant, W.R. (1956): On the Measurement of Prevalence of Arthritis and Rheumatism from Interview Data. *Journal of Chronic Diseases*, 3, pp. 134–139.
- Coder, J.F. (1980): Some Results from the 1979 ISDP Research Panel. *American Statistical Association, Proceedings of the Section on Survey Research Methods*, pp. 540–545.
- Commission on Chronic Illness (1957): *Chronic Illness in a Large City: The Baltimore Study*. Harvard University Press, Cambridge, MA.
- Cowan, C.D., Roman, A.M., Wolter, K.M., and Woltman, H.F. (1979): A Test of Data Collection Methodologies: The Methods Test. *American Statistical Association, Proceedings of the Section on Survey Research Methods*, pp. 141–146.
- Deighton, R.E. (1967): Some Results of Experimentation with Self-Respondent Interviewing Procedures, February 1965 – June 1966. U.S. Bureau of the Census memorandum (February 28, 1967), Washington, D.C.
- Elinson, J. and Trussell, R.E. (1957): Some Factors Relating to Degree of Correspondence for Diagnostic Information as Obtained by Household Interviews and Clinical Examinations. *American Journal of Public Health*, 47, pp. 311–321.
- Enterline, P.E. and Capt, K.G. (1959): A Validation of Information Provided by Household Respondents in Health Surveys. *American Journal of Public Health*, 49, pp. 205–212.
- Ferber, R. (1955a): On the Reliability of Purchase Influence Studies. *Journal of Marketing*, 19, pp. 225–232.
- Ferber, R. (1955b): On the Reliability of Responses Secured in Sample Surveys. *Journal of the American Statistical Association*, 50, pp. 788–810.
- Ferber, R. and Frankel, M. (1981): Evaluation of the Reliability of the Net Worth Data in the 1979 Panel: Asset Ownership on Wave 1. University of Illinois, Survey Research Laboratory, Champaign-Urbana, IL.
- Haase, K.E. and Wilson, R.W. (1972): The Study Design of an Experiment to Measure the Effects of Using Proxy Respondents in the National Health Interview Survey. *American Statistical Association, Proceedings of the Social Statistics Section*, pp. 289–293.
- Haberman, P.W. and Elinson, J. (1967): Family Income Reported in Surveys: Husbands Versus Wives. *Journal of Marketing Research*, 4, pp. 191–194.
- Hansen, M.H., Hurwitz, W.N., Marks, E.S., and Mauldin, W.P. (1951): Response Errors in Surveys. *Journal of the American Statistical Association*, 46, pp. 147–190.
- Hill, D.H. (1987): Assessing the Relative Quality of Proxy Data from a Non-Experimental Study. Paper presented at the Third Annual Research Conference (March, 1987), U.S. Bureau of the Census, Washington, D.C.
- Horvitz, D.G. (1952): Sampling and Field Procedures of the Pittsburgh Morbidity Survey. *Public Health Reports*, 67, pp. 1003–1012.
- Jones, C. and Aquilino, R. (1970): Methods Test Phase III: Second Report on the Accuracy of Retrospective Interviewing and Effects of Nonself Response Labor Force Status. U.S. Bureau of the Census memorandum (January 29, 1970), Washington, D.C.
- Kaluzny, R.L. (1981a): Evaluation of Experi-



- mental Effects, 1979 Research Panel – Wave 1. Mathematica Policy Research, Inc., Princeton, NJ.
- Kaluzny, R.L. (1981b): Evaluation of Experimental Effects, 1979 Research Panel – Wave 2. Mathematica Policy Research, Inc., Princeton, NJ.
- Kilss, B. and Alvey, W. (1976): Further Exploration of CPS-IRS-SSA Wage Reporting Differences for 1972. American Statistical Association, Proceedings of the Social Statistics Section, pp. 471–476.
- Kinsey, A.C., Pomeroy, W.B., and Martin, C.E. (1948): Sexual Behavior in the Human Male. Saunders, Inc., Philadelphia, PA.
- Kohn, M.L. and Carroll, E.E. (1960): Social Class and the Allocation of Parental Responsibilities. Sociometry, 23, pp. 372–392.
- Kolomel, L.N., Hirohula, T., and Nomura, A. (1977): Adequacy of Survey Data Collected from Substitute Respondents. American Journal of Epidemiology, 106, pp. 476–484.
- Koons, D.A. (1973): Quality Control and Measurement of Nonsampling Error in the Health Interview Survey. National Center for Health Statistics Pub. No. (HSM) 73–1328, Rockville, MD.
- Kovar, M.G. and Wilson, R.W. (1976): Perceived Health Status – How Good is Proxy Reporting? American Statistical Association, Proceedings of the Social Statistics Section, pp. 495–500.
- Kovar, M.G. and Wright, R.W. (1973): An Experiment with Alternate Respondent Rules in the National Health Interview Survey. American Statistical Association, Proceedings of the Social Statistics Section, pp. 311–316.
- Krueger, D.E. (1957): Measurement of Prevalence of Chronic Disease by Household Interviews and Clinical Evaluation. American Journal of Public Health, 47, pp. 953–960.
- Kulka, R.A. (1982): Tests and Experiments. Chapter 4 in Research Triangle Institute, ISDP 1979 Research Panel Documentation, Research Triangle Institute, Research Triangle Park, NC.
- Levinger, G. (1966): Systematic Distortion of Spouses' Reports of Preferred and Actual Sexual Behavior. Sociometry, 29, pp. 291–299.
- Marquis, K.H. (1983): Record Checks for Sample Surveys. Paper presented at the Advanced Research Seminar on Cognitive Aspects of Survey Methodology (June, 1983), Committee on National Statistics, National Research Council, Washington, D.C.
- Marquis, K.H., Duan, N., Marquis, M.S., and Polich, J.M. (1981): Response Errors in Sensitive Topic Surveys. Rand Corporation, Santa Monica, CA.
- Martin, E.A. and Biderman, A.D. (1984): Memorandum on Proxy vs. Self-response for 12–13 Year-olds. Crime Survey Research Consortium Items 101–141, Bureau of Social Science Research, Washington, D.C.
- Martin, J. and Butcher, B. (1982): The Quality of Proxy Information – Some Results from a Large-scale Study. The Statistician, 31, pp. 293–319.
- Mathiowetz, N. and Groves, R.M. (1983): The Effects of Respondent Rules on Health Survey Reports. Chapter V in C.F. Cannell et al., An Experimental Comparison of Telephone and Personal Health Surveys, National Center for Health Statistics, Rockville, MD.
- Mooney, H.W. (1962): Methodology in Two California Health Surveys: San Jose (1952) and Statewide (1954–55). Public Health Monograph No. 70, Government Printing Office, Washington, D.C.
- Mudd, E.H., Stein, M., and Mitchell, H.E. (1961): Paired Reports of Sexual Behavior of Husbands and Wives in Conflicted Marriages. Comprehensive Psychiatry, 2, pp. 149–156.
- Neter, J. and Waksberg, J. (1965): Response Errors in Collection of Expenditures Data

- by Household Interviews: An Experimental Study. Technical Paper No. 11, U.S. Bureau of the Census, Washington, D.C.
- Nisselson, H. and Woolsey, T.D. (1959): Some Problems of the Household Interview Design for the National Health Survey. *Journal of the American Statistical Association*, 54, pp. 69–87.
- Olson, J. (1981): Self-Proxy Treatment Rules and Panel Attrition. U.S. Department of Health and Human Services memorandum (April 20, 1981), Washington, D.C.
- O'Muircheartaigh, C. (1986): Correlates of Reinterview Response Inconsistency in the Current Population Survey (CPS). Paper presented at the Second Annual Research Conference (March 1986), U.S. Bureau of the Census, Washington, D.C.
- Radke, M.J. (1946): *The Relation of Parental Authority to Children's Behavior and Attitudes*. University of Minnesota Press, Minneapolis, MN.
- Roman, A.M. (1983): Preliminary Findings from Comparisons of Self Versus Proxy Response for Students from the 1979 ISDP. U.S. Bureau of the Census memorandum (September 29, 1983), Washington, D.C.
- Roman, A.M. and Woltman, H.F. (1980): The Methods Test Panel: Analysis of the Unemployment Rate: An Interim Report: June 1978 – March 1979. U.S. Bureau of the Census report (March 14, 1980), Washington, D.C.
- Roshwalb, A. (1982): Respondent Selection Procedures Within Households. *American Statistical Association, Proceedings of the Section on Survey Research Methods*, pp. 93–98.
- Rutter, M. and Brown, G.W. (1966): The Reliability and Validity of Measures of Family Life and Relationships in Families Containing a Psychiatric Patient. *Social Psychiatry*, 1, pp. 38–53.
- Singh, M.P. and Tessier, R. (1975): MTP (Methods Test Panel) Phase III: Feasibility Test for Complete Non-proxy Procedure, Methodology and Analysis. Statistics Canada, Ottawa, Canada.
- Sudman, S. and Bradburn, N.M. (1974): *Response Effects in Surveys*. Aldine, Chicago, IL.
- Thompson, D.J. and Tauber, J. (1957): Household Survey, Individual Interview, and Clinical Examination to Determine Prevalence of Heart Disease. *American Journal of Public Health*, 47, pp. 1131–1140.
- Turner, A.G. (1972): Methodological Issues in the Development of the National Crime Survey Panel: Partial Findings. Law Enforcement Assistance Association unpublished report (December 1972), Washington, D.C.
- Turner, C.F. and Martin, E.A. (1984): *Surveying Subjective Phenomena*. Basic Books, New York, NY.
- Vaughan, D. (1980): Preliminary Findings on the Data Quality Effects of Employing Self-response Rules in the 1979 ISDP Panel. Social Security Administration draft memorandum (February 16, 1980), Washington, D.C.
- Vaughan, D. (undated a): Interview Visits by Response Mode and Form Type, February Interview, 1979 ISDP Panel. Social Security Administration draft memorandum, Washington, D.C.
- Vaughan, D. (undated b): Some Evidence on Response Mode and Refusal Rates, Wave One and Two of the 1979 ISDP Panel. Social Security Administration draft memorandum, Washington, D.C.
- Williams, L.E. (1969): Methods Test Phase III: First Report on the Accuracy of Retrospective Interviewing and Effects of Non-self Response on Labor Force Status. U.S. Bureau of the Census memorandum (June 24, 1969), Washington, D.C.