Latent Class Analysis of Response Inconsistency across Modes in NSFG Cycle 6

Ting Yan[1], Roger Tourangeau[1,2], Frauke Kreuter[2]
Social Research Center, University of Michigan[1]
Joint Program in Survey Methodology, University of Maryland[2]

Oct. 12, 2006
Draft. Paper To Be Presented at the 2006 National Survey of Family Growth Research Conference. Please don't cite without contacting the authors first.

Abstract

Latent class analysis (LCA) has been used to model measurement error, to identify flawed survey questions, and to compare mode effects. Using data from NSFG Cycle 6, we examined the measurement error properties of responses to questions asked in two different modes (CAPI and ACASI). To evaluate mode effects within LCA strong assumptions need to be made about prevalence and response error rates in subdomains of the population. This paper will demonstrate such an analysis and discuss the "surprising" finding of high false positive and false negatives rates in ACASI compared to CAPI.
Latent Class Analysis, Mode Comparison, NSFG, CAPI, ACASI, False Positive, False Negative

1 Introduction

Most surveys rely on self-reports to collect data on sexual behaviors. Self-report is, therefore, an integral component in estimating the prevalence of contraceptive use, the number of sexual partners, or cohabitation histories, and the prevalence of abortion or miscarriage and so on. The consensus in the survey literature is that self-report is susceptible to measurement error from numerous sources: question wording and question order (Schuman and Presser, 1981; Schwarz et al., 1991; Tourangeau et al., 1991), misunderstanding of survey concepts (Conrad and Schober, 2000; Schober and Conrad, 1997), recalling error (Neter and Waksberg, 1964), inappropriate use of estimation strategies (Conrad et al., 1998), deliberate misreporting (Tourangeau and Smith, 1996), and the effects of the interview mode (Tourangeau and Smith, 1996). However, the size and direction of measurement error associated with self-report are often unknown.

For sensitive topics (such as sexual behavior), studies suggest that the presence of an interviewer yields underreporting of socially undesirable behavior and likewise overreporting of socially desirable behavior (Tourangeau et al., 2000). The mode effect literature indicates that self-administered modes such as ACASI can increase reports of socially undesirable behavior (Turner et al., 1998). This is often interpreted as a reduction in underreporting - that is, a reduction in the false negative rate relative to interviewer-administered interviews.

In the absence of true scores, the examination of error reduction relies on a 'more-is-better' assumption. For example, for questions about socially undesirable behaviors such as drug use or having had an abortion, the assumption is that the mode of interview that produces a higher level of reports of the sensitive behavior is less biased. There are at least two problems with this approach.

- First, even if the 'more-is-better' assumption can be used to identify which mode of data collection is less biased, it is still unknown how much better the better mode is.
- Second, the 'more-is-better' assumption may not be valid for all segments of the population. Tourangeau and Smith (1996) found, for instance, that unmarried males tend to exaggerate their sexual activity while unmarried females tend to underreport theirs (see Smith 1996 for a review of similar findings). The more-is-better assumption would be misleading for the males in this case.

The present study uses the latent class analysis (LCA) approach to estimate measurement biases for questions about sexual behavior[1]. The major advantage of LCA is that it does not require true values or gold standards to produce estimates of false negative and false positive rates. However, it does require multiple measurements of the same underlying construct (or the latent variable); the multiple measurements can be either similar questions embedded in the same survey or repeated measures in different waves of a survey.

In the National Survey of Family Growth (Cycle 6), a number of questions on sensitive topics are asked twice during the interview. They are asked by an interviewer in computer assisted personal interviews (CAPI) and filled out by respondents through audio computer-assisted self-interviewing (ACASI). Thus, the two different modes of administration contribute two indicators getting at the same underlying concept. Following existing examples in the survey literature (Biemer and Witt, 1996; Sinclair and Gastwirth, 1996; Biemer et al., 2001; Biemer and Bushery, 1999; Biemer, 2001; Biemer and Wiesen, 2002;

[1]This study is based upon work supported by the National Science Foundation under Grant No.0550002. We like to thank the Methodology, Measurement, and Statistics (MMS) Program and Dr. Eavey for their support. Any opinions, findings, and conclusions or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of the National Science Foundation. The authors contributed equally to this paper.
Contact address: tingyan@isr.umich.edu

Biemer, 2004a,b) we proposed to use LCA to contribute to the following research questions:

What are the estimated error rates by modes, using LCA? Does ACASI always yield better data than CAPI? And if yes, how much better?

Given LCA's heavy reliance on assumptions to make models identifiable, we will also discuss the appropriateness and usefulness of using LCA in the given context. How strong do the assumptions have to be with only two indicators? Do the LCA results look plausible? And what can be learned from the LCA results?

2 Latent Class Analysis

The standard latent class analysis measures one or more unobserved (latent) categorical variables through a set of observed indicators. The basic idea of LCA is that the associations between the observed indicators can be explained by the latent variable(s). The relationship between the indicators is attributed to the varying distribution of indicators across different classes of the latent variable. Within each latent class, the observed variables are unrelated. This 'local independence' (Lazarsfeld and Henry, 1968) assumption allows inferences about the latent class variable.

For a set of binary indicators, the relationship between the indicator and the latent variable can be described with logistic regression equations. In such models, the probability for each observed item u (from a set of J observed items) is the product of the probability of being in class k of the latent class variable c and the probability of the observed response ($u_j = 1$), given class membership, summed across all of the latent classes:

$$P(u_j = 1) = \sum_{k=1}^{K} P(c = k)P(u_j = 1|c = k)$$

Following the notation of Muthén (2001), the joint probability of all J observed variables (u_1, u_2, etc.) under the assumption of conditional independence is

$$P(U_i = u_i) = \sum_{k=1}^{K} \gamma_k \prod_{i=1}^{I} \rho_{i|k}.$$

LCA produces unconditional probabilities $\gamma_k = P(c = k)$, which represents the probability that respondents are assigned to each class of the latent variable. In a way, the unconditional probabilities estimate the prevalence of each class in the population (or the size of each latent class). In addition, LCA also allows one to obtain various probabilities conditional on class membership. For example in a two class model the probability to endorse a binary item u_1 conditioned on being in class one will be estimated $\rho_{1|1} = P(u_1 = 1|c = 1)$ and the probability to not endorse this particular item $\rho_{2|1} = P(u_1 = 2|c = 1)$; similarly for class two $\rho_{1|2} = P(u_1 = 1|c = 2)$ and $\rho_{2|2} = P(u_1 = 2|c = 2)$. Two of the conditional probabilities represent the extent of misclassifications produced by the survey questions; they are the probability of a false positive response and the probability of a false negative response for a question item given membership in the latent class. These are sometimes referred to collectively as the "error probabilities". A high false positive probability or a high false negative probability usually signals that there might be a problem with a particular survey question. Thus, LCA allows comparisons of question sensitivity and specificity to identify questions that best differentiate the classes.

2.1 Hui-Walter model assumptions

Mathematically speaking, three binary indicator variables are needed for a model of *two* latent classes to be just identified. However, in survey context, three measurements don't come by easily. When there are only two indicators, one can impose assumptions on the parameters ρ to achieve identifiability.

For instance, one possible assumption restricts the false negative probability to zero or sets the latent class size to be equal across classes[2]. However, sometimes these assumptions are either too stringent or theoretically implausible. Another way to free up additional degrees of freedom is to use a covariate $g_i = 1, 2, \ldots, G$ predicting latent class membership to inform the model. A grouping variable can be added to establish identifiability if restrictions on γ or ρ are imposed. Biemer and Witt (1996) refer to it as the *Hui-Walter* model (Hui and Walter, 1980). The grouping variable has to satisfy two assumptions:

1. The prevalence rates have to be different across the levels of the grouping variable (the different prevalence assumption).

2. The false positive and false negative probabilities have to be equal across levels of the grouping variable (the equal error probabilities assumption).

One would, for example, assume different prevalence of each labor force category (employed, unemployed) for males and females $P(c = 1|g = 1) \neq P(c = 1|g = 2)$, but males and females would have the same probability to endorse the item given their true state (class membership) $\rho_{1|1,g=1} = \rho_{1|1,g=2}$. Or to phrase this differently, there is no reason to assume that males and females make gender specific errors in answering the employment question.

2.2 Applications

Biemer (2004a) used the Hui-Walter model to estimate error in labor force question in the 1996 CPS. Using data from the original survey and a reinterview, Biemer fit a LCA model without constraining the error probabilities for the interview and reinterview to be identical. The estimated misclassification probabilities showed a high misclassification rate for the unemployed status; according to

[2]For examples of various restrictive or equality assumptions, see McCutcheon (1987)

the results, about one third of unemployed persons were misclassified in the CPS and 80% of the misclassified cases were incorrectly classified as "not in labor force" (Biemer, 2004a). This finding was consistent with both historical data on the reliability of the CPS data and theoretical expectations suggesting that the concept of unemployment is a difficult one for many respondents (Biemer, 2004a,b).

In another study Biemer and Witt (1996) examined data from the 1993-1994 NHSDA, focusing on reports about the use of three drugs alcohol, marijuana, and cocaine. For each drug, an LCA model with two indicators was fit to the data. One indicator was based on responses to a question asking the respondents about the most recent time they used a particular drug. The other indicator was a composite based on a series of questions about each drug. The respondents status as a smoker or nonsmoker was used to partition the population into two domains, as required by the Hui-Walter procedure. Biemer and Witt compared the estimated prevalence rate obtained from the recency question alone and from the Hui-Walter models with the NHSDA estimator. They found a consistent pattern in the prevalence estimates that is, recency question resulted in smaller prevalence rates than the LCA estimators, which are in turn smaller than the NHSDA estimators. This was because the LCA estimators using the Hui-Walter model were adjusted for false positive and false negative errors, while the NHSDA estimator ignored these errors. In a similar vein, Biemer (2001) used LCA to estimate the measurement bias in CATI and face-to-face components of the NHIS survey. He found that the face-to-face interviews contained larger measurement bias than the CATI interviews. His analyses also provided evidence that measurement bias can exceed nonresponse bias for many items in a survey even when response rates are relatively low.

For the present paper we will use the same set of assumptions as was used by Biemer and colleagues. We will use two questions that are asked both in the CAPI and ACASI portion of the interview as the indicators and then a covariate to inform the latent class model. We will choose a covariate for which we can assume different prevalence rates and equal error rates.

3 Data

NSFG Cycle 6 was completed in 2002; it was based on an area probability sample that represented the civilian noninstitutionalized population of the United States, 15-44 years of age. The Survey Research Center of the University of Michigan carried out the data collection, completing in-person interviews with 12,571 respondents - 7,643 females and 4,928 males. The overall response rate was 79 percent. The NSFG questionnaire covers several sensitive topics, including the number of sexual partners the respondents have had, their fertility and cohabitation histories, their use of contraception, and their history of sexually transmitted diseases.

A unique feature of the NSFG interview is that it uses two modes of data collection. In Cycle 6, the early sections of the interview were done via CAPI; one of the later sections was done via ACASI. With ACASI, the computer displays the questions on-screen directly to the respondents while playing a recording of the questions to them via earphones. Several studies have demonstrated the advantages of ACASI over interviewer-administered questionnaires for the collection of sensitive information (e.g., Tourangeau and Smith, 1996; Turner et al., 1998).

In Cycle 6 of the NSFG, several items were administered in both sections of the questionnaire. For example, the ACASI section of the female questionnaire asked "Has a male ever put his penis in your vagina (also known as vaginal intercourse)?" The CAPI section includes a similar item: "At any time in your life, have you ever had sexual intercourse with a man, that is, made love, had sex, or gone all the way?" There are several other variables that can be derived both from responses to the CAPI section of the interview and from responses to the ACASI section. For example, the ACASI questionnaire asks directly "How many pregnancies have you had that ended in abortion?" In CAPI, respondents are asked the following question: "In which of the ways shown on Card 12 did the pregnancy end?"[3]

3.1 Indicator Variables

In this paper, we focus on two questions: whether respondents ever had an abortion (or ever made someone have an abortion) and whether they ever had sex. Both pieces of information are sensitive in nature, but they are subject to reporting errors of different direction. The level of perceived sensitivity and the likelihood for misreporting is likely to depend on the cultural context and respondents attitudes towards the issues. For (most) females in the American society, admitting having had an abortion is embarrassing, but admitting ever having had sex is not. In other words, one would expect to see underreports on having an abortion and overreports on having sex (men are especially subject to overreporting sexual behaviors). Thus, the abortion question is more likely to attract false negative responses whereas the sex question false positive responses. We chose these two variables on purpose to examine whether LCA is robust and sensitive enough to detect errors of opposite directions.

3.2 Selection of Grouping Variables

The NSFG data only allows two-indicator latent class models. In order to make the models identifiable, we need grouping variables to inform the models. However, the grouping variables have to satisfy the Hui-Walter assumptions of different prevalence rates and equal error

[3]Where Card 12 showed Miscarriage, Stillbirth, Abortion, Ectopic or tubal pregnancy, Live birth by Cesarean section, Live birth by vaginal delivery. This question is repeated for all pregnancies respondents reported. An abortion variable is computed based on responses to this question.

rates. The different prevalence rates for certain subgroups of the population can be established using external data sources. Ideally, several data sources are to be used to identify appropriate grouping variables. The use of multiple data sources is to reduce the interference of measurement error associated with each source. If we can establish a pattern of different prevalence groups across sources, we feel reasonably confident that those serve well as grouping variables.

To satisfy the assumption of equal error rates, subgroups need to be found that have similar probabilities to over- or underreport in a given data collection mode. One factor assumed to be related to error rates is the perceived sensitivity of a question. Or to put it differently the attitude a respondent has toward the survey topic.

The NSFG includes items assessing attitudes towards premarital sex, traditional sex roles, and other potentially relevant issues that might predict whether respondents hold liberal or traditional attitudes or values about life. We assume this difference in ideology and attitudes in turn affect the error probabilities. Thus, we divide the sample into groups with similar attitudes and values (i.e., liberal women and traditional women), assuming error probabilities are likely to be similar within each group. Latent class models are conducted for people holding liberal attitudes and traditional attitudes separately. Martial status (ever married vs never married) was used as the grouping variable to inform the latent class models. It seemed reasonable to assume similar error rates for the abortion item among those who have ever been married and those who never married once we control for differences in attitudes[4]. The assumptions about the error rates in the "ever sex" item are less clear. Hypotheses can be formed about different error rates across marital status for the "every sex" item, an issue we will get back to in the discussion.

4 Analyses and Results

4.1 Abortion

4.1.1 Female Respondents

Our first set of analyses focus on abortion. Table 1 presents the prevalence of females who ever had an abortion using CAPI and ACASI. It is apparent from the table that ACASI produced higher prevalence of abortion than CAPI for traditional and liberal women.

If the "more-is-better" assumption were employed to evaluate data quality by mode, the conclusion would be that ACASI estimate is less biased than the CAPI estimate. However, this simplistic conclusion wouldn't shed light on whether the less bias is due to a higher false positive rate associated with the ACASI responses or a higher false negative associated with the CAPI responses. Furthermore, what remains unknown is the extent of underreporting (or the size of the false negative responses) and of overreporting (i.e., the false positive rates) in CAPI and in ACASI.

[4]A full set of 13 attitudinal items is used to form an additive liberalism index. The exact wordings are listed in the Appendix.

Traditional Women		Liberal Women	
CAPI	ACASI	CAPI	ACASI
19.8	23.1	31.5	34.1

Table 1: Estimated Prevalence (in %) of People Who Ever Had an Abortion From Female Survey

Mode	Prevalence Estimate	False Positive	False Negative
CAPI	19.8	0.005	0.133
ACASI	23.1	0.008	0.000
LCA	23.1		

Table 2: Estimated Error Rates from LCA Models, Female Survey (Traditional)

Looking at the subset of traditional women first (see Table 2), we see in line with previous studies higher false negative rates (more denials) in the interviewer-administered mode. To phrase it differently, the results suggest that women underreport more in CAPI than in ACASI. If we are confident that the Hui-Walter assumptions are met, the LCA results seem to suggest that both modes suffer from some false positives as well, with a slightly higer rate in ACASI than in CAPI. Nevertheless, the LCA estimate of prevalence rate for traditional women that matches the ACASI estimate.

Next we examined the liberal women (see Table 3). As expected, we see lower false negative rates in both modes compared to the traditional women. However, given that the Hui-Walter assumptions are met, this analysis suggests that ACASI results could be affected by both false negative and false positive rates. LCA results also suggest that, for the liberal women, the comparatively large false positive response led to an ACASI estimate higher than CAPI. The LCA estimate matches, here, the CAPI prevalence rate.

4.1.2 Male Respondents

We examined a similar question asked in the male survey. Table 4 displays the estimated prevalence rates of abortion obtained from CAPI and ACASI. Similar to the female results, the ACASI prevalence estimates are consistently higher than the CAPI estimates for traditional

Mode	Prevalence Estimate	False Positive	False Negative
CAPI	31.5	0.000	0.003
ACASI	34.1	0.043	0.015
LCA	31.5		

Table 3: Estimated Error Rates from LCA Models, Female Survey (Liberal)

CAPI	ACASI	CAPI	ACASI
20.4	28.5	24.4	32.7

Table 4: Prevalence Estimates (in %) of People Who Ever Had an Abortion, Male Survey Traditional Men Liberal Men

Mode	Prevalence Estimate	False Positive	False Negative
CAPI	20.4	0.011	0.187
ACASI	28.5	0.059	0.002
LCA	22.7		

Table 5: Estimated Error Rates from LCA Models, Male Survey (Traditional)

and liberal men.

In line with previous research, LCA confirms higher false negative rates for both modes in both subgroups (traditional and liberal men). However, the LCA estimates showed larger false positive rates in ACASI for traditional men, leading to an overall LCA estimate of abortion that is between the CAPi and ACASI estimate.

For the subgroup of liberal men (Table 6), no false positive rates are estimated in either one of the two modes. Here again LCA estimates larger false negative rates in CAPI compared to ACASI. The overall estimate of the LCA is close to the ACASI estimate.

4.2 Ever Sex

4.2.1 Female Respondents

Our next set of analyses examines responses to the "ever sex" question. The estimated prevalence rates are similar across modes and across the two women groups, as shown in Table 7. The CAPI estimates are slightly higher than the ACASI estimates for both groups (χ^2=8.22, $p < .01$ for traditional women, (χ^2=.47, *n.s.* for liberal women). Since overreporting (rather than underreporting) is the threat to data quality in this case, the "more-is-better" evaluation criterion is not applicable. Still, a direct comparison by mode seemed to suggest that ACASI is better than CAPI in reducing overreporting of ever having had sex; in other words, there seemed to be fewer false positive responses in ACASI than in CAPI.

However, the error rates obtained from the latent class approach, displayed in Table 8, tell a different story. Contrary to expectation, the false positive responses were

Mode	Prevalence Estimate	False Positive	False Negative
CAPI	24.4	0.000	0.259
ACASI	32.7	0.000	0.007
LCA	32.9		

Table 6: Estimated Error Rates from LCA Models, Male Survey (Liberal)

Traditional Women			Liberal Women		
CAPI	ACASI	LCA	CAPI	ACASI	LCA
88.5	86.2	88.5	89.3	88.8	89.3

Table 7: Estimated Prevalence (in %) Estimated Prevalence of Females Who Ever Had Sex, Female Survey

	Traditional Women		Liberal Women	
Mode	FP	FN	FP	FN
CAPI	0.038	0.000	0.000	0.000
ACASI	0.058	0.028	0.110	0.019

Table 8: Estimated Error Rates from LCA Models, Female Survey

higher in ACASI than in CAPI for women. Furthermore, the ACASI responses also showed a 2-3% of false negative responses. Thus, the ACASI estimates were worse than the CAPI estimates, contradictory to what was suggested by the direct comparisons of the two mode estimates. At least for female data, the CAPI estimates are closer to the LCA estimates than the ACASI estimates, which were subject to both false positive and false negative error.

4.2.2 Male Respondents

Results from the male data are quite comparable to those from the female data. The ACASI estimates of prevalence of people who ever had sex are slightly lower than the CAPI estimates (see Table 9) (χ^2=2.19, *n.s.* for traditional men, (χ^2=2.67, *n.s.* for liberal men). However, the error rates estimates from the LCA approach demonstrated that the ACASI data showed higher false positive responses than the CAPI data. In addition, the ACASI data is not free from false negative responses as the CAPI data did. Therefore, the LCA results seemed to suggest that ACASI didn't offer data of better quality than CAPI given the presence of both false positive and false negative responses.

4.3 Discussion

This paper investigates the possibility of using latent class analysis to examine mode effects. In the absence of true values (or "gold standards"), latent class approach is an appealing alternative to evaluate data quality associated with different modes. The LCA approach does not require true values or error-free indicators. All it needs are two or more indicator questions that get at the same information. In return, the LCA produces an

Traditional Men			Liberal Men		
CAPI	ACASI	LCA	CAPI	ACASI	LCA
81.9	80.2	81.9	84.8	83.1	84.8

Table 9: Prevalence Estimates (in %) Estimated Prevalence of People Who Ever Had Sex, Male Survey

	Traditional Men		Liberal Men	
Mode	FP	FN	FP	FN
CAPI	0.000	0.000	0.014	0.000
ACASI	0.047	0.032	0.056	0.028

Table 10: Estimated Error Rates to the Sex Question from LCA Models, Male Survey

estimate of the overall prevalence rate and quantitative estimates of the error rates (false positive and false negative rates). However, LCA d relies on a different set of assumptions that involve error rates and prevalence rates for the grouping variables.

In this paper, we examined responses to two sensitive questions ("ever had an abortion" or "ever made some one have an abortion" and "ever had sex"). Both questions are asked in the CAPI and ACASI components of the survey. We used the CAPI and ACASI answers as two indicators that were later entered into various latent class models. The results from the latent class analysis provided detailed and potentially useful information for mode comparison. Even though the ACASI estimates of abortion prevalence are higher than the CAPI estimates for both females and males, we found that the ACASI estimates are not always bias free. If the assumptions used in our LCA models hold, the traditional 'more-is-better' assumption would yield misleading conclusions regarding the quality of ACASI data for traditional men and women.

The LCA results also showed that the estimated error rates differed across modes. However, at this point, One can only speculate on the sources of these errors. The presence of false positives in ACASI could, for example, be a function of response burden, low motivation, or fatigue on respondents' part, given that ACASI questions were asked at the end of a long CAPI interview. The finding that more false negative responses are present in CAPI data than in ACASI is consistent with survey literature on mode effects. The only exception is for the liberal women, who had relatively smaller false negatives in CAPI than in ACASI.

ACASI responses to the sex questions manifested higher false positive rates than the CAPI responses, contrary to the common expectation that ACASI elicits more honest responses. For the "ever sex" item, the LCA results indicated that the false negative rates are zero for CAPI responses, identifying CAPI as a better mode than ACASI in producing data of high quality. This seemingly implausible result should warn us about the vulnerability of LCA model and prompt us to cast doubts on our model specification.

Even though the LCA approach is able to produce more detailed information on error rates, we acknowledge the limitations with the LCA approach. A key limitation lies in its heavy reliance on the Hui-Walter assumptions when a grouping variable is used to make the model identifiable. The validity of the LCA results depends on whether the two assumptions are satisfied or not; thus, the choice of a grouping variable is critical. The use of a grouping variable that violates either the different-prevalence-rate assumption or the equal-error-rate assumption could produce biased results and misleading conclusions. We assumed in this paper that people holding similar attitudes or values are likely to have similar response behavior, yielding similar error rates within groups. External data source Laumann et al. (1994) showed that people who ever married have a different prevalence rate in abortion and having had sex from those who never married. Thus, we assume that our grouping variable satisfies the Hui-Walter assumptions. However, it is open for discussion whether these assumptions are any stronger or weaker than the "more-is-better' assumption usually applied in mode comparison studies. Our next step is to examine the robustness of the LCA approach with grouping variables violating the Hui-Walter assumptions to different extent. In addition, we will consider applying other identifying restrictions to achieve identifiability. For instance, we could set the false positive rates for the abortion item to zero since overreporting on abortion is very rare. We could also set the prevalence rate of "every had sex" to one for married respondents.

To compare data quality across modes, it is not enough to just point out which mode is subject to a higher false positive response and which mode is to higher false negative response. Practitioners and researchers also need a calibration method to adjust for estimates obtained from the inferior mode. We will address this issue by examining the robustness of the prevalence estimates provided by LCA. We will further study LCA and its use in mode comparison.

4.4 Appendix

Next, I would like to get your opinion on some matters concerning family life. I will read you some statements, and I would like you to tell me if you strongly agree, agree, disagree, or strongly disagree. The first is:

- It is better for a person to get married than to go through life being single. Do you strongly agree, agree, disagree, or strongly disagree?
- Divorce is usually the best solution when a couple can't seem to work out their marriage problems.
- Sexual relations between two adults of the same sex are all right. Do you strongly agree, agree, disagree, or strongly disagree?
- Any sexual act between two consenting adults is all right.
- It is all right for unmarried 18 year olds to have sexual intercourse if they have strong affection for each other.
- It is all right for unmarried 16 year olds to have sexual intercourse if they have strong affection for each other.
- The rewards of being a parent are worth it, despite the cost and the work it takes.
- It is okay for an unmarried female to have a child. Would you say you strongly agree, agree, disagree, or strongly disagree?)

- Gay or lesbian adults should have the right to adopt children.
- A young couple should not live together unless they are married.
- A working mother can establish just as warm and secure a relationship with her children as a mother who does not work.
- It is much better for everyone if the man earns the main living and the woman takes care of the home and family.
- It is more important for a man to spend a lot of time with his family than to be successful at his career. Do you strongly agree, agree, disagree, or strongly disagree?

References

Biemer, P. (2001). Nonresponse bias and measurement bias in a comparison of face to face and telephone interviewing. *Journal of Official Statistics 17*, 295–320.

Biemer, P. and J. Bushery (1999). Estimating the error in labor force data using markov latent class analysis. *Proceedings of the FCSM Conference*.

Biemer, P. and C. Wiesen (2002). Measurement error evaluation of self-reported drug use: A latent class analysis of the us national household survey on drug abuse. *Journal of Royal Statistics 165*, 97–119.

Biemer, P. P. (2004a). An analysis of classification error for the revised current population survey employment questions. *Survey Methodology 30*, 127–140.

Biemer, P. P. (2004b). Modeling measurement error to identify flawed questions. In S. e. a. Presser (Ed.), *Methods for Testing and Evaluating Survey Questionnaires*, pp. 225–246. New York: Wiley.

Biemer, P. P. and M. Witt (1996). Estimation of measurement bias in self-reports of drug use with applications to the national household survey on drug abuse. *Journal of Official Statistics 12*, 275–300.

Biemer, P. P., H. Woltman, D. Raglin, and J. Hill (2001). Enumeration accuracy in a population census: An evaluation using latent class analysis. *Journal of Official Statistics 17*, 129–148.

Conrad, F. G., N. R. Brown, and E. R. Cashman (1998). Strategies for estimating behavioral frequency in survey interviews. *Memory 6*, 339–366.

Conrad, F. G. and M. Schober (2000). Clarifying question meaning in a household telephone survey. *Public Opinion Quarterly 64*, 1–28.

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

Laumann, E. O., J. H. Gagnon, R. T. Michael, and S. Michaels (1994). *The Social Organization of Sexuality: Sexual Practices in the United States*. The University of Chicago Press.

Lazarsfeld, P. F. and N. W. Henry (1968). *Latent Structure Analysis*. Boston: Houghton Mifflin.

McCutcheon, A. L. (1987). *Latent class analysis*. Beverly Hills: Sage Publications.

Muthén, B. (2001). Second-generation structural equation modelling with a combination of categorical and continuous latent variables: New opportunities for latent class/latent growth modelling. In A. Sayer and L. Collins (Eds.), *New methods for the analysis of change*, pp. 291–322. Washington, DC: American Psychological Association.

Neter, J. and J. Waksberg (1964). A study of response errors in expenditures data from household interviews. *Journal of the American Statistical Association 59*, 17–55.

Schober, M. F. and F. G. Conrad (1997). Does conversational interviewing reduce survey measurement error? *Public Opinion Quarterly 61*, 576–602.

Schuman, H. and S. Presser (1981). *Questions and answers in attitude surveys*. New York: Academic Press.

Schwarz, N., F. Strack, and H. P. Mai (1991). Assimilation and contrast effects in part-whole question sequences: A conversational logic analysis. *Public Opinion Quarterly 55*, 3–23.

Sinclair, M. D. and J. L. Gastwirth (1996). On procedures for evaluating the effectiveness of reinterview survey methods: Application to labor force data. *JJournal of the American Statistical Assocation 91*, 961–969.

Tourangeau, R., K. Rasinski, and N. Bradburn (1991). Measuring happiness in surveys: A test of the subtraction hypothesis. *Public Opinion Quarterly 55*, 255–266.

Tourangeau, R., L. Rips, and K. Rasinski (2000). *The psychology of survey response*. Cambridge: Cambridge University Press.

Tourangeau, R. and T. W. Smith (1996). Asking sensitive questions: The impact of data collection mode, question format, and question context. *Public Opinion Quarterly 60*, 275–304.

Turner, C. F., L. Ku, S. M. Rogers, J. H. Lindberg, L. D.and Pleck, and F. L. Sonenstein (1998). Adolescent sexual behavior, drug use, and violence: Increased reporting with computer survey technology. *Science 280*, 867–873.